Peter Chapman · Gem Stapleton
Amirouche Moktefi · Sarah Perez-Kriz
Francesco Bellucci (Eds.)

LNAI 10871

# Diagrammatic Representation and Inference

## 10th International Conference, Diagrams 2018
## Edinburgh, UK, June 18–22, 2018
## Proceedings

Springer

# Lecture Notes in Artificial Intelligence 10871

Subseries of Lecture Notes in Computer Science

LNAI Series Editors

Randy Goebel
*University of Alberta, Edmonton, Canada*
Yuzuru Tanaka
*Hokkaido University, Sapporo, Japan*
Wolfgang Wahlster
*DFKI and Saarland University, Saarbrücken, Germany*

LNAI Founding Series Editor

Joerg Siekmann
*DFKI and Saarland University, Saarbrücken, Germany*

More information about this series at http://www.springer.com/series/1244

Peter Chapman · Gem Stapleton
Amirouche Moktefi · Sarah Perez-Kriz
Francesco Bellucci (Eds.)

# Diagrammatic Representation and Inference

## 10th International Conference, Diagrams 2018
## Edinburgh, UK, June 18–22, 2018
## Proceedings

Springer

*Editors*
Peter Chapman
Edinburgh Napier University
Edinburgh
UK

Gem Stapleton
University of Brighton
Brighton
UK

Amirouche Moktefi
Tallinn University of Technology
Tallinn
Estonia

Sarah Perez-Kriz
Mitre Corporation
McLean, VA
USA

Francesco Bellucci
University of Bologna
Bologna
Italy

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-319-91375-9 ISBN 978-3-319-91376-6 (eBook)
https://doi.org/10.1007/978-3-319-91376-6

Library of Congress Control Number: 2018942344

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer International Publishing AG, part of Springer Nature 2018, corrected publication 2019
Chapter "How Cross-Representational Signaling Affects Learning from Text and Picture: An Eye-Tracking Study" is Open Access. This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). For further details see license information in the chapter.
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG
part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

The 10th International Conference on the Theory and Application of Diagrams (Diagrams 2018) was hosted by Edinburgh Napier University during June 18–22, 2018. For the first time, Diagrams was co-located with the 23rd International Conference on Conceptual Structures (ICCS), which stimulated discussion between researchers of both communities.

Submissions to Diagrams 2018 were solicited in the form of long papers, short papers, posters and, for the first time in the conference's history, Abstracts. The Abstracts category was introduced to encourage participation from authors working in fields where publications at conferences are not as prestigious as those appearing in journals.

The peer-review process involved all papers and abstracts receiving at least three reviews (at least two for posters) by members of the Program Committee or a nominated sub reviewer. After reviews were received, authors had the opportunity to submit a rebuttal. The reviews and rebuttals led to a lively discussion involving the Program Committee and the conference chairs to ensure the highest-quality submissions, covering a broad range of topics, were accepted for the conference.

We would like to thank the Program Committee members and the additional reviewers for their considerable contributions. The robust review process, in which they were so engaged, is a crucial part of delivering a major conference. In total, 124 submissions were received. Of these, 26 were accepted as long papers. A further 28 were accepted as short papers, 20 as Abstracts, and 20 for poster presentation.

Diagrams 2018 sought to expand the research community, and introduced two special submission tracks on Philosophy and Psychology after noting their underrepresentation in past editions of the conference. This action was extremely successful with 20, out of the total 124, submissions made to the Psychology track and 54 submissions to the Philosophy track. In addition, 28 Abstracts were submitted across the main track and the two special tracks, with strong representation from both philosophers and psychologists.

Alongside the technical program, Diagrams 2018 included a Graduate Symposium, the Set Visualization and Reasoning Workshop, and six tutorials covering a diverse set of topics on diagrams. There were two keynote speakers. Professor Ahti-Veikko Pietarinen, from Tallinn University of Technology, gave an excellent keynote talk on philosophical aspects of diagrams. Professor Keith Stenning, from the University of Edinburgh, gave a joint keynote with ICCS covering research on diagrams from a cognitive perspective.

There are, of course, many people to whom we are indebted for their considerable assistance in making Diagrams 2018 a success. We thank Sarah Perez-Kriz, for her role as Abstracts Chair, and Renata de Freitas, for her role as Workshops and Tutorials Chair. We are grateful to Andrew Blake for organizing the Graduate Symposium, which supported the next generation of researchers. Atsushi Shimojima did an outstanding job as Publicity Chair, no doubt reflected by the number of submissions. Francesco Bellucci took on the demanding role of Proceedings Chair, for which we are immensely grateful. We are grateful to the NSF for providing significant funding, which supported graduate students from US-based institutions with their participation. Our institutions, Edinburgh Napier University, the University of Brighton, and Tallinn University of Technology, also provided support for our participation, for which we are grateful. Amirouche Moktefi was further supported by the ERC project “Abduction in the Age of Uncertainty” (PUT 1305, Principal Investigator: Prof. Ahti-Veikko Pietarinen). Lastly, we thank the Diagrams Steering Committee for their continual support, advice, and encouragement.

June 2018

Peter Chapman
Gem Stapleton
Amirouche Moktefi

# Organization

## Program Committee

| | |
|---|---|
| Francesco Bellucci | University of Bologna, Italy |
| Lisa Best | University of New Brunswick, Canada |
| Mireille Betrancourt | University of Geneva, Switzerland |
| Alan Blackwell | University of Cambridge, UK |
| Jean-Michel Boucheix | LEAD-CNRS, University of Burgundy, France |
| Richard Burns | West Chester University, USA |
| Jim Burton | University of Brighton, UK |
| Jean-Yves Béziau | University of Brazil, Brazil |
| Jessica Carter | University of Southern Denmark, Denmark |
| B. Chandrasekaran | The Ohio State University, USA |
| Peter Chapman | Edinburgh Napier University, UK |
| Peter Cheng | University of Sussex, UK |
| Jean-Marie Chevalier | Collège de France, France |
| Daniele Chiffi | Tallinn University of Technology, Estonia |
| Lopamudra Choudhury | Jadavpur University, India |
| Peter Coppin | OCAD University, Canada |
| James Corter | Columbia University, USA |
| Frithjof Dau | SAP, Germany |
| Jim Davies | Carleton University, Canada |
| Silvia De Toffoli | Stanford University, USA |
| Aidan Delaney | University of Brighton, UK |
| Lorenz Demey | K.U. Leuven, Belgium |
| Maria Giulia Dondero | Fonds National de la Recherche Scientifique/Université de Liège, Belgium |
| Tim Dwyer | Monash University, Australia |
| George Englebretsen | Bishop's University, Canada |
| Isobel Falconer | University of St. Andrews, UK |
| Jacques Fleuriot | The University of Edinburgh, UK |
| Janet Folina | Macalester College, USA |
| Valeria Giardino | Archives Henri Poincaré, Nancy, France |
| Emily Grosholz | The Pennsylvania State University, USA |
| Mary Hegarty | University of California, Santa Barbara, USA |
| John Howse | University of Brighton, UK |
| Mateja Jamnik | University of Cambridge, UK |
| Mikkel Willum Johansen | University of Copenhagen, Denmark |
| Yasuhiro Katagiri | Future University Hakodate, Japan |
| Sarah Kriz | The MITRE Corporation, USA |

| | |
|---|---|
| John Kulvicki | Dartmouth College, USA |
| Brendan Larvor | University of Hertfordshire, UK |
| John Lee | The University of Edinburgh, UK |
| Catherine Legg | Deakin University, Australia |
| Javier Legris | Universidad de Buenos Aires, Argentina |
| Jens Lemanski | FernUniversität in Hagen, Germany |
| Christina Ljungberg | University of Zurich, Switzerland |
| Danielle Macbeth | Haverford College, USA |
| Emmanuel Manalo | Kyoto University, Japan |
| Kim Marriott | Monash University, Australia |
| Luana Micallef | Helsinki Institute for Information Technology HIIT, Finland |
| Amirouche Moktefi | Tallinn University of Technology, Estonia |
| Hari Narayanan | Auburn University, USA |
| Marco Panza | CNRS, France |
| Ahti-Veikko Pietarinen | Tallinn University of Technology, Estonia |
| Beryl Plimmer | The University of Auckland, New Zealand |
| Helen Purchase | University of Glasgow, UK |
| Peter Rodgers | University of Kent, UK |
| Yuri Sato | University of Brighton, UK |
| Dirk Schlimm | McGill University, Canada |
| Stephanie Schwartz | Millersville University, USA |
| Atsushi Shimojima | Doshisha University, Japan |
| Sun-Joo Shin | Yale University, USA |
| Jurgis Skilters | CCSS, Germany |
| Hans Smessaert | K.U. Leuven, Belgium |
| Gem Stapleton | University of Brighton, UK |
| Ben Steichen | California State Polytechnic University, Pomona, USA |
| Frederick Stjernfelt | Aalborg University, Denmark |
| Andrew Stull | University of California Santa Barbara, USA |
| Takeshi Sugio | Doshisha University, Japan |
| Ryo Takemura | Nihon University, Japan |
| Yuri Uesaka | The University of Tokyo, Japan |
| Jean Paul van Bendegem | Vrije Universiteit Brussel, Belgium |
| Agnes Veszelszki | Corvinus University of Budapest, Hungary |
| Petrucio Viana | Federal Fluminense University, Brazil |

## Additional Reviewers

Bach, Benjamin
Blake, Andrew
Chakraborty, Mihir
Chang, Chunlei
Hullman, Jessica
Klein, Karsten
Linker, Sven
Morris, Imogen
Nakagawa, Masanori
Shams, Zohreh
Sim, Kevin
Smeltzer, Karl

# Contents

## Keynote Contributions

## Tutorials

## Generating and Drawing Euler Diagrams

## Diagrams in Mathematics

## Diagram Design, Principles and Classification

## Reasoning with Diagrams

## Euler and Venn Diagrams

## Empirical Studies and Cognition

## Peirce and Existential Graphs

## Logic and Diagrams

## Posters

# Keynote Contributions

# Diagrams and Nonmonotonic Logic: What Is the Cognitive Relation?

Keith Stenning(✉)

University of Edinburgh, Edinburgh, UK
k.stenning@ed.ac.uk

[9] summarised a program of research into diagrammatic reasoning based on the semantics of diagrams. It distinguished *direct* diagrams from *indirect* ones by whether there was a direct semantic interpretation of some spatial relations in the diagram or whether all the interpreted spatial properties were mediated by a concatenation relation, as is the case in written natural language. So an Euler diagram is direct because its spatial property of containment in a labelled closed curve is directly interpreted as membership in a set designated by the label. In a typical abstract network diagram, such as a semantic graph, the nodes-and-link configurations take the place of the concatenation operator (as well as other matters), and play similar roles in determining the semantics of the diagram indirectly as the concatenation operator does in written natural and logical languages. The hallmark of indirectness is that the concatenation relation (whether one- or two-dimensional) has itself no semantic interpretation—only a syntactic one.

Directness gives rise to specific *inexpressiveness* of direct diagrammatic systems, and while indirect diagrammatic systems may be inexpressive, the limitation on that expressiveness is not a result of the properties of the 'spatial concatenation operator'. The program's argument was that this limitation of expressiveness played important roles in determining the cognitive properties of direct diagrammatic systems. Inexpressiveness generally leads to tractability of reasoning (in the computer science sense), and tractability of reasoning generally leads to ease of cognitive use. One kind of example is that the reasoning to an interpretation of direct diagrams *may* be simple, and gaining a grip on the interpretation may be nine-tenths of the problem naive users have with some direct diagrammatic systems.

This ease of interpretation is possibly what gives rise to the widespread illusion that there is no need to interpret diagrams: they provide their interpretation 'on their face'. This may be the case in examples where the interpretation is 'culturally available', but is in general not true, even in very simple cases such as Euler's diagrams. Notably, these diagrams were widely misinterpreted by a range of psychologists, leading to 'demonstrations' that these diagrams could not provide the psychological basis for syllogism solution. Euler had a specific strategy for the use of his diagrams that remained largely implicit in [3], but which is absolutely essential to the diagram's advantage in tractability of reasoning. With the flat-footed interpretation proposed by the psychologists (the one that provides the proof that the diagrams are not used mentally), the reasoning becomes

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 3–8, 2018.
https://doi.org/10.1007/978-3-319-91376-6_1

quite intractable, at least for a reasoner with the architecture of human working memory. For example, the syllogism *Some A are B. Some B are C. Some A are C* requires consideration of about 60 candidate models, and an auxilliary search for propositions true in all of them, of which, of course, there are none. With more careful experimentation, and a one-slide introduction to how to interpret the diagrams (what Euler does in his letters to the Princess), the diagrams are highly beneficial, because they induce the understanding of the syllogism required, an interpretation that few undergraduate subjects faced with the usual task adopt.

This approach to diagrams was mainly explored empirically in the teaching of logic more generally, through a collaboration with Jon Barwise and John Etchemendy using their computer environment for heterogeneous reasoning to teach first order logic *Hyperproof*. The most striking empirical finding of this program was the discovery of deep individual differences in the ways that undergraduates on their first logic course used the diagrams, with resulting improvements in their 'general reasoning'. Apart that is, from the demonstration that real logic courses can make large improvements in the reasoning of even highly selected undergraduates, an idea that was, and still is, routinely dismissed in psychology, on little evidence. For those interested in the teaching or instruction of diagrammatic reasoning, these results are probably the most important take-home. For the 'weak expressiveness' approach, they were a complication. It looked superficially, as if the theory worked for half the class. It forced more care in the statement of the theory. *Interpretation is everything* might be a slogan for the change of emphasis. To understand the effects of direct diagrams required painstaking attention to the details of the interpretations that they induced: not a bad slogan for the psychology of reasoning generally?

This talk is about what happened next. Diagrams originally caught my attention while I was, as usual, researching natural language discourse. They offered a counterpoint to language. Being told that using logical analysis of natural language discourse means the psychology is limited to linguistic reasoning had grown tedious. My first love had been the idea that there was an alternative logic for the construction of models for narratives. The counterpoint of diagrams reinforced the idea that interpretation had to be central. But in natural language narrative processing, the involvement of interpretation is far more all encompassing. Unlike the situation in classical logic where interpretation is a starter for the real meal, in discourse processing, interpretation is the main course. The problem had been in the early years [8], that classical logic was obviously the wrong logic, but there wasn't as yet a suitable alternative. Nonmonotonic logics had been invented [4] but it wasn't easy to see how they could be applied. Although they had been invented to make 'ordinary' reasoning with general knowledge easy, it had turned out that their tractability was even worse than classical logic, and here it wasn't clear that they could be fitted to small problems, because they could immediately explode by demanding the retrieval of just about any piece of human knowledge to connect the first two sentences of a discourse.

At this point, *Seeing Reason* was just being shepherded to the publisher, when I met Michiel van Lambalgen in Amsterdam. Michiel is a logician and a probabalist, and a Kant scholar. A logician was what I needed for the nonmonotonic logic; a probabilist for dealing with what had become the opposition; and Kant was my first philosophical hero. More immediately relevant was that Michiel announced that he wanted to understand the experimental approach to cognitive issues. I quoted my exalted supervisor's reply to a related question I had asked him years earlier: "When I get an idea for an experiment nowadays, I lie down in a dark room and wait for it to go away" was his reply. This didn't seem to put Michiel off. From Michiel I learned how much logic had happened while I did my diagrams-blink. Now there was *preferred model semantics* for nonmonotonic logics [7], and PROLOG (which I only knew as a programming language) had been recognised as a tractable nonmonotonic logic [2]. With these two, one had a hope of explaining how natural language narrative was, in its logical properties, nearly the opposite of the classical logic with which I, and more importantly Grice, had been inappropriately struggling to understand cooperative communication. But even more than the tools, Michiel (collaborating with Fritz Hamm) [12], had shown that Constraint Logic Programming plus Kowalski's Event Calculus [6] could give a nonmonotonic semantics for natural language narrative. Matters were just waiting for a psychologist to do a lot of learning.

So the real purpose of this talk to this audience is to explore the similarities between the multiple-logics program that grew out our encounter, and the 'weakly expressive direct diagrams' work already introduced. Logic Programming (LP), the nonmonotonic logic that grew out of PROLOG, is 'weak' in a number of ways. It's birth gives the most general hint. How did PROLOG arise? From the gleam in the eye that logic ought to be usable for programming computers at a much higher level than the then available imperative languages, combined with the ghastly realisation that classical logic was fundamentally intractable. Much of our understanding of computation came from the study of the *un*decidability of classical logic. So LP is based on the 'Horn-clause fragment' of classical logic, which was the most promising fragment that is tractable. So weakness was the fundamental desideratum (coupled of course with just enough expressiveness for a programming language).

LP has a conditional with an *abnormality* clause $ab$, which is negated. So $(p \wedge \neg ab) \rightarrow q$ is read as "If p, and nothing is abnormal, then q". This makes LP's conditional *exception tolerant*. If, in the context of processing, something is abnormal, then $\neg ab$ becomes false, and the antecedent conjunction $p \wedge \neg ab$ also becomes false, which blocks the inference to $q$. How does one know there is an abnormality? Well, there is an abnormality list: a disjunction of exceptions, the truth of any one of which makes $ab$ (the abstract abnormality proposition) true.

It is therefore important to understand the status of such a conditional in this case. It does *not* become false: it becomes inapplicable. These conditionals are organised into potentially very large collections (knowledgebases (KBs)) (think human semantic memory) which can be expressed as feedforward networks [10]. A retrieval in the task of answering a query leads to flow through this feedforward network (analogous to psychology's informal concept of 'spreading activation').

In fact, an implementation of procedural memory such as ACT-R [1] is close to an LP network, though its applications are different from ours. So if there is evidence that *ab* is true (one of the disjoined list of particular abnormalities is true) then the truth of $p$ will not trigger the conclusion $q$. That is all. The conditional remains in place. If the evidence of abnormality is subsequently overridden then the conditional will fire in successive retrievals. In fact, the conditionals function more like contentful inference rules than the propositions of classical logic. To accommodate this behaviour, a 3-valued Kleene semantics is used in which the 'intermediate' truth-value **I** is read 'indeterminate' rather than the 'intermediate' familiar from say Lukasiewicz logic. Indeterminate is needed because the vast majority of conditionals in a KB are inactive during any episode of reasoning.

The existence of a KB means that search for propositions can have a much stronger interpretation than in classical logic. Not finding $p$ as a premiss just means you can't use $p$ in inferences. In LP, *failure-to-find* $p$ can be interpreted as establishing that it is true that $\neg p$. This inference is called 'negation-as-failure'. If a discourse starts: "Once upon a time there was a cat", then a search for giraffes, at this point, will yield the truth of "There is no giraffe". So spreading activation search through the LP-net yields all the propositions entailed by the givens in the current model, and the current model is updated by the addition of any new ones. Despite the technical complexities needed to achieve it, LP has the property that a unique minimal model (the *preferred* model) results at every addition of a proposition to the input. This should be immediately attractive to psychologists who know the psycholinguistic literature on peoples' construction of models for discourses. If we cannot get a single model of the gist of a story, we typically feel we have no understanding, and demonstrably can remember little of the sentences that made it up.

So the differences with classical logic are stark. Instead of infinitely many models existing for many sets of premisses, we get a single model for any coherent pair of KBs. In classical logic the conclusions can contain nothing new (nothing that wasn't implicit in the premisses). In LP every sentence needs to add something new (and maybe subtract things) to/from the preferred model, and so the interpretation changes at every point. This is cooperative reasoning *to* an interpretation (a unique minimal preferred model) in counterpoint to classical logic's adversarial reasoning *from* an interpretation *to* a propositional conclusion.

LP is nomonotonic as opposed to classical logic's monotonicity, and whatever is in common between Euler's circles reasoning to a classical logical syllogism conclusions, and LP reasoning to a preferred model of a story, the logics are different. We need multiple logics for the incompatible things that people do in reasoning. The answer lies in interpretation. To understand the process of arriving at an interpretation of Euler diagrams (the one we need to do syllogisms easily), then we need LP to model that process. If we look at the 'one slide tutorial' that explains Euler's tactics/strategy for using his circles to at least a good proportion of our subjects [5], then we could implement it in LP, though to give a serious implementation we'd need to do some empirical work on what other interpretations there are, which one is the one our students jump to, and so on. The output of that process would be an account of the interpretative process in this particular situation with these particular students, which would output a model of Euler's interpretation which would constitute the decision. Even once a subject has that interpretation, there may be considerable tactical skill needed in applying it to difficult syllogisms. In classical logic there is a separable *theorem prover* which tells you which proof rules to apply at each point. This calculation and the skill of applying it, are separable from the conceptual knowledge of validity embodied in the logic. [11] shows that there is at least one algorithm which is abstract with regard to whether it is implemented in a sentential logical account, or a diagrammatic Euler account, and which can explain how people can do the reasoning once the interpretation is understood. Many psychological experiments have proceeded on the assumption that the reasoning has to be different on the two implementations. Different it is in the superficial detail, but the same it is when one gets below to the logical core.

So LP and Euler's circles are utterly different even though they are both chosen for their weak expressivity. But LP's design to output unique minimal preferred models, rather than single sentence conclusions, and the enforcement of the provision of information by direct diagrams such as Euler's circles share this feature. The talk will close with some expansion of this point. In classical logic, logicians define models simply as sets of sentences. And any consistent set of sentences is a model of some set of premisses. But in LP, and more fastidiously, Constraint LP + the Event Calculus, there are certain 'connectedness' and 'coherence' constraints that have to be applied to filter out sets of sentences which have no models, despite being consistent. So, in narrative, there are constraints that certain temporal, spatial, referential and identity information is provided, that generally is not all explicitly in the discourse. One can see these constraints operating in simple examples like *Max fell. John pushed him.* In classical logic, in the obvious interpretation, these two sentences express propositions that are logically unrelated. As initial in a story context, they typically induce an interpretation in which Max's pushing causes John's falling, and therefore precedes it. Neither sentence alone entails this information: only as a discourse are they consequences in CLP, and contained in its preferred model, at this point. They may of course nonmonotonically disappear with certain continuations. This required coherence and completeness of models (as opposed to incoherent or incomplete sets of sentences) is like the completeness of specification of sets by the primitive topological structure of the circles in an Euler diagram that does not permit any vagueness about which types are represented as possible by any of the permissible diagrams. Euler's 'theorem prover' which is then laid on top of this minimal diagrammatic structure, circumvents this enforcement of the representation of all possible types, and eases the identification of conclusions. And the structure can be a useful aid to teaching the ideas that 'searching all possible types' is necessary for Euler's theorem prover to work by exploiting the weakness of syllogistic semantics: any valid syllogism can be modelled with a model containing only a single element. So weakness of expression does turn out to be an important commonality, and to have implications for the practical uses of diagrams, if only one gets the comparison at the right levels. However trivially simple are Euler's diagrammatic method and the syllogisms it was invented for, they provide a structure for the study of at least direct diagrammatic systems of a wide variety.

## References

1. Anderson, J.R.: The Architecture of Cognition. Psychology Press, Hove (2013)
2. Doets, K.: From Logic to Logic Programming. MIT Press, Cambridge (1994)
3. Euler, L.: Lettres à une princesse d'Allemagne: sur divers sujets de physique & de philosophie. PPUR presses polytechniques, Lausanne (2003)
4. McCarthy, J.: Circumscription-a form of non-monotonic reasoning. In: Readings in Artificial Intelligence, pp. 466–472. Elsevier (1981)
5. Monaghan, P., Stenning, K.: Generalising individual differences and strategies across different deductive reasoning domains. In: Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making, p. 45. Wiley, Hoboken (2004)
6. Shanahan, M.: The event calculus explained. In: Wooldridge, M.J., Veloso, M. (eds.) Artificial Intelligence Today. LNCS (LNAI), vol. 1600, pp. 409–430. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48317-9_17
7. Shoham, Y.: A semantical approach to non-monotonic logics. In: Proceedings of Tenth International Joint Conference on Artificial Intelligence, pp. 388–392 (1987)
8. Stenning, K.: Anaphora as an approach to pragmatics. MIT Press, Cambridge (1978)
9. Stenning, K.: Seeing Reason: Image and Language in Learning to Think. Oxford University Press, Oxford (2002)
10. Stenning, K., Van Lambalgen, M.: Human Reasoning and Cognitive Science. MIT Press, Cambridge (2012)
11. Stenning, K., Yule, P.: Image and language in human reasoning: a syllogistic illustration. Cognit. Psychol. **34**(2), 109–159 (1997)
12. Van Lambalgen, M., Hamm, F.: The Proper Treatment of Events, vol. 6. Wiley, Hoboken (2008)

# The Beauty of Graphs

Ahti-Veikko Pietarinen[1,2](✉)

[1] Tallinn University of Technology, Tallinn, Estonia
`ahti-veikko.pietarinen@ttu.ee`
[2] Nazarbayev University, Astana, Kazakhstan

**Abstract.** The beauty of logical graphs consists in many facets, including notational simplicity, multi-modality and normativity. This paper aims at understanding the nature of Peirce's graphical method and its implications to philosophy of logic.

**Keywords:** Logical graphs · Existential graphs · Beauty · Logicality

## 1 Introduction

Adapting Russell's words on the beauty of mathematics, a study of Peirce's method leaves little doubt that "a logical graph, rightly viewed, possesses not only truth, but supreme beauty". This beauty has not merely or even predominantly an aesthetic but an exact logical and intellectual quality. The continuity of the sheet on which logical graphs are scribed makes the sheet in Peirce's terms a "perfect sign" (R 283(s)). Being perfect, it has got to be the "quasi-mind" that unifies symbol, index and icon. Graphs are propositional symbols ("phemes"), and thus picture one or more of the three categories: intellectual concepts, thoughts, or generalities.

Our exploration on the beauty of graphs comes in three parts. First, theories of logical graphs can exploit *notational simplicity*. Graphical representations of logical constants perform well when they unify the notational apparatus of the theory. With such notational simplicity, other things then follow. Notational advances contribute to the development of science. Changes—sometimes radical—in notations facilitate discovery and innovation. Advances in science are invariably preceded by representational advances in systems of signs as much as they are in methods of communication. Obviously notations and communication methods co-evolve.

The concept of notational parsimony is itself not a simple one. The movement of a double pendulum is described by unassuming formulas whose predictions are tedious to compute. A three-body system is simple but has no formalization that we could solve out without infinities. The two are very different cases. Peirce's

---

A. -V. Pietarinen—Supported by the Estonian Research Council's Personal Research Grant PUT 1305 and Nazarbayev University's Social Policy Grant 2018–2019. My thanks to Francesco Bellucci for comments.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 9–12, 2018.
https://doi.org/10.1007/978-3-319-91376-6_2

higher-order graphs that represent anteriority and posteriority relations are large and occupy several bridges, but the meaning of these relations is not complicated. They are pictures of generality. Yet the logic consisting of such graphs would be incomplete and undecidable. To describe a concept by measuring the simplicity of graphs by the number and size of the areas it needs to enclose, plus the number of non-connected lines that cross each other, we can get one measure of simplicity; this has no correlation with the simplicity in the alphabet that uses only few, and minimally just one, notational primitives.

Notation that simplifies representation of logical constanthood yields advances in the following sense. When the same sign, such as an oval, signifies both the scope (parentheses) and an operation, the system gains in *analyticity*. The precise notion of analyticity is crucial. Graphs in the Beta notation use lines that abut ovals. Disconnected and resting on different areas with loose ends they increase the quantificational depth of the graph. Graphs in the Alpha notation consist of nested ovals that represent both the scope of logical constants and one-place operations such as negation or involution. There may or may not be a sign for the conditional. Illation suggests that there is.

Alternatively, nests or the scroll may mean the presence only of an implication, thus analyzing the relation of illation in its ultimate details. The smallest axiom for implicational graph calculus emerges from insertions and iterations and has 13 areas, yielding syntactic completeness. The freedom one gains is an increase in analytic virtue: these minutest steps of logical reasoning are suspended well in the graphical method.

Second, the beauty of graphs lies in the fact that they are *multi-modal*. They need not appeal to the eye, not even predominantly so. It may be misleading to think logical graphs as unique because of their allegedly superior ocularity over other logical notations [6]. Notations aid the course of interpretation of the relations of its parts. Those relations have to be perceived and conceived in some sensible way. For example, a diagrammatic Alpha can work with sounds, polyphony and silence, where the sheet is white noise [5].

Conventions of a tactile logic can appeal to a variety of sensations. Ambient space could be that of texture, density, state, depth, temperature or vibration. Values could be represented by haptic, cutaneous and muscular polarities between rough and smooth, hard and soft, dry and wet, and so on. The degrees of freedom are immense, with practical thrust [2]. Peirce's *tinctured graphs* were a preliminary exploration of these unusual representational spaces for various sensations, moods and modalities.

Perception of beauty is also a *moral experience*. If graphs make a good instrument of logical analysis of meaning and mind, this would be a great benefit in value. No child currently reads a diagram unless trained to do so. Does it presuppose linguistic competence? At least rhematic, skeletal level comprehension may be needed. But if graphs can help discern truth better than other systems, would we have a duty to teach them to the young? Do they do that? There seems to be a gain also in synthetic virtues. We can run a Kantian thought-experiment which takes graphs as objects of reasoning and construe imagined objects (now taking graphs as interpolants) for a particularly non-analytic example of quantificational reasoning. Our observation of those constructions in the imagination has immediately shown the validity of this example. It does so without recourse to complex diagrammatic constructions that interpret objects in the universe of discourse. Phenomenality of transformations enlightens logicality.

All three faces of beauty emerge from certain fundamental departures from the conventions of traditional logical notation. The three most important ones are the following. First, one begins with the *sheet*. This ambient space is a continuous manifold upon which graphs are scribed. Yet the blank is itself also an assertion in the language of graphs and represents the graph of coexistence.

The uses and significations of the sheet can vary. They make it evident that different logical behaviours will ensue. Peirce used the sheet to signify not only truth but assertion, proof, possibility, the future, the destined, and even imperative and interrogative moods. We can choose from his palette, say, epistemic, temporal, provability or illocutionary graphs and develop them further than he managed to do.

In the other direction, we can disrobe the sheet of certain meanings. We can conceive that it cannot be cut, for example. We will get systems that no longer entertain certain basic logical truths. Either way, the sheet motivates the emergence of wildly different systems.

Second, transformations exploit *deep inference*. This falls from the dimensionality of the sheet. As it is a feature lacking in the way ordinary notations are set up, attempts to improve those notations have not yet come to terms with the full meaning of the depth of proofs. Since applying permissive moves in graphs means that we can dive deep into their areas, the graphical notation may solve problems that have troubled commonplace inferential systems.

Third, there is unity and continuity in the two perfect signs of system: the sheet and the lines of identity. The lines show the connectivity of areas and bonding of spots. This leads to the supposition that Wittgenstein was indeed right: notions of the quantifier and identity should not have gotten separated in the first place. We could do well to eliminate equality from logical notations and express identity of the object by identity of the sign, not by the sign of identity.

If asserting identity of two objects is in Wittgenstein's words "nonsense", self-identity is a pseudo-graphical form like an empty cut is a non-asserted boundary or symmetry and transitivity rules that "cannot even be written down" (Tr. 5.534). Self-identity is in Beta shown as a thick loop of the line of identity. Thus well-defined objects are represented not by extremities of lines but by loops or "swellings" (R 493) at their ends. In addition to the blank, the line is in Peirce's theory that other sign which is both continuous and perfect.

Several further elements partake of this constitution of beauty. That continuous predicates are *indecomposable*, that there is a fundamental role for graphs of *teridentity* in the method, and the fact that Peirce's theory of logical graphs is the only known method in which *logical* and *material* (topological) *continuities* are strictly correlated [1], contribute to the purely logical beauty of graphs.

The sheet itself is internal to the language of graphs, and can be a representation of all truths (or assertions, transformations), or else a representation of everything being false (or denials, disproofs). The former emerged in late 1896 as "Positive Logical Graphs" (R 488). Shortly afterwards, Peirce renamed them *existential graphs*, to emphasize assertions of facts that positively exist in the universe. The latter, namely the case where the blank means falsity, are the *entitative graphs*. They give rise to the system of contradictions.

Ultimately Peirce did not think that either is the proper language of graphs. Philosophical (logical) and psychological (phenomenological) considerations are values complementary to each other. This gives rise to the *Existential–Entitative Duality*: a function $f$ mapping existential to entitative graphs and $g$ mapping entitative to existential graphs form a Galois connection. Indeed Peirce wrote in his long Christmas gift to James that logic analyzes the generalization of the "direct perception of what we are immediately aware of" (R L 224, December 25, 1909). In this letter Peirce predicts his analytic method to be "the Logic of the Future" [4].

The leverage of graphs on the philosophy of logic is high. As scope and duality belong together, so does quantification and identity. The very idea of the leading principle of reasoning (Peirce's Rule a.k.a. residuation) arises from observing certain areas of logical diagrams spread on the sheet as blank. The problem of justifying deductive reasoning resolves itself at once. The gap between vision and reason is fulfilled. One might hope that the graphical method provides unity and overarching analogies to discussions on logical pluralism when they become occupied with things like axioms, proofs and computations. Going spatial harbors non-trivial results and drives conceptual change. Many mathematicians try to find adequate geometrical languages for physics.

Last, I conjecture that logical graphs excite those Brodmann areas that are responsible for emotional experiences, whereas traditional (non-graphical) logical notations may fail to do so. The two do not share the same experiences of beauty. To see the beauty of graphs is to see the change in view; the change is to see it as a theory that is simple, homogenous and unusually uberous.

## References

1. Bellucci, F.: Peirce's continuous predicates. Trans. Charles S. Peirce Soc. A Quart. J. Am. Philos. **49**(2), 178–202 (2013)
2. Pagliano, P.: The Multisensory Handbook: A Guide for Children and Adults with Sensory Learning Disabilities. Routledge, London (2012)
3. Peirce, C.S.: Manuscripts. Houghton Library of Harvard University. Cited as R followed by manuscript number (1939–1914)
4. Peirce, C.S.: Logic of the Future: Writings on Existential Graphs. Ed. by Pietarinen, A.-V. (2018, to appear)
5. Pietarinen, A.-V.: Is non-visual diagrammatic logic possible? In: Gerner, A. (ed.) Diagrammatology and Diagram Praxis. College Publications, London (2010)
6. Pietarinen, A.-V., Bellucci, F.: Two dogmas of diagrammatic reasoning: a view from existential graphs. In: Hull, K., Atkins, R.K. (eds.) Peirce on Perception and Reasoning: From Icons to Logic, pp. 174–195. Routledge, London (2017)

# Tutorials

# Were "Super-Turing" Diagrammatic Reasoning Mechanisms Ancient Products of Biological Evolution?

Aaron Sloman(✉)

The University of Birmingham, Birmingham, UK
`A.Sloman@cs.bham.ac.uk`
`http://www.cs.bham.ac.uk/~axs`

**Abstract.** Immanuel Kant had understood some deep facts about mathematical knowledge that have mostly been ignored by recent researchers on cognition, especially intrinsic connections between many everyday actions and mathematical competences, e.g. competences concerned with the fact that organisms inhabit environments with complex mathematical structures, some produced by activities of life forms, others not. I'll present a variety of examples, to be discussed with the audience, with deep implications for future research in artificial and natural cognition, and raise questions about the diagram-like information structures many cognitive processes seem to be concerned with: "diagrams in the mind". I suspect that Alan Turing's 1952 paper on chemical morphogenesis, published two years before he died is connected with this. Perhaps if he had lived several more decades he would have worked on what I call the *Meta-Morphogenesis* project, which was inspired by Turing and has deep connections with mathematical structures important for animal cognition and future machine cognition. However it is an open question that current forms of (digital) computation will need to be enhanced using chemistry-based computation similar to sub-neural mechanisms in brains, or whether the required forms of reasoning can occur in virtual machines implemented in digital machinery. Von Neumann's last little book, written as he was dying in 1958, raises similar questions.

**Keywords:** Evolution of ancient spatial mathematical competences
Evolved construction kits: concrete and abstract
Biological foundations for mathematics
Evolution: the blind mathematician
Forms of representation for intelligent machines
Examples of toddler and non-human mathematical competences

## 1 Introduction

I believe that by 1781 Immanuel Kant had understood some deep facts about mathematical knowledge that had been implicitly denied by David Hume, and

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 15–19, 2018.
https://doi.org/10.1007/978-3-319-91376-6_3

have mostly been ignored by recent researchers, especially intrinsic connections between many everyday actions and mathematical competences. Examples include competences concerned with the fact that organisms inhabit environments with complex and varied mathematical structures. Some of the structures are produced by activities of life forms, such as honeycombs, spider-webs, and many plant forms, whereas others arise out of inanimate processes, including solar and lunar effects, weather, volcanoes, erosion, rainfall, wave actions, etc.

I shall use much of the tutorial to present examples for discussion, in the hope that some of those attending will have important new ideas. Anyone thinking of attending can find out about some of the examples from a collection of freely available online web pages produced over several years, and recently incorporated into the framework of the Turing inspired Meta-Morphogenesis project, presented here: http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html.

Participants will be invited to comment on several types of example relevant to this project, some of them already presented online, e.g. in

http://www.cs.bham.ac.uk/research/projects/cogaff/misc/toddler-theorems.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/triangle-sum.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/torus.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/deform-triangle.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/rubber-bands.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/trisect.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/cardinal-ordinal-numbers.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/cup-saucer-challenge.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/shirt.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/rings.html
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/cardinal-ordinal-numbers.html.

In 1969 McCarthy and Hayes [1] published a logicist manifesto for AI, suggesting that intelligent machines (and by implication humans and other animals) could in principle perform all required information processing using logical forms of representation, based entirely on Fregean forms of representation (using only formalisms based on functions applied to arguments).

I criticised this at IJCAI in [2] suggesting that what I (misleadingly) called “analogical” representations, in which properties and relations represent properties and relations (often in context-sensitive ways that rule out isomorphism with what is represented) could often have greater heuristic power than Fregean representations. Similar ideas have been reinvented several times, especially in the context of research on diagrammatic reasoning.

Since the 1980s, there has been increasing interest in rival mechanisms based on neurally inspired models of computation (e.g. [3]) given an enormous boost by advances in speed and reductions in price of computers (a process limited by the end of Moore's law). Rival waves of enthusiasm have been concerned with embodied cognition, situated cognition, and no doubt several others, often apparently based on personal preferences and a tiny sample of examples, rather than careful analysis of requirements.

I suggest that we should try to learn from biological evolution: what were the challenges and opportunities faced at various stages of evolution, and in various kinds of environment, and how did genetic changes arise in response. This requires very careful comparative analysis of requirements for organisms with different capabilities and environments with different opportunities and challenges. This is the goal of the Meta-Morphogenesis project.

The tutorial will present aspects of this project, including the importance for evolution of the concept of a "construction kit". The physical/chemical universe provides a *fundamental* construction kit (FCK), but biological evolution created and used increasingly varied and complex *derived* construction kits (DCKs) of many forms, discussed in this long, incomplete, highly speculative paper (an earlier version of which was published in 2017): http://www.cs.bham.ac.uk/research/projects/cogaff/misc/construction-kits.html.

This approach criticises not only the purely logicist approach to AI, but also reliance on neural theories that are inadequate because they cannot explain mathematical discoveries (e.g. by Archimedes, Euclid, Zeno, etc.). Related theories over-emphasising embodied cognition also fail to address some of the deepest aspects of perception, learning, and development, especially the growth of different competence layers based on genetic mechanisms that evolved at different times and instead of all being expressed at birth are "staggered" in ways proposed in collaboration with biologist Chappell and Sloman [4]. See also http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-configured-genome.html.

Some of the facts about geometrical and topological competences based on understanding mathematical impossibilities and necessary connections also seem to challenge current models of deep learning and current theories of brain function neither of which seems able to explain how impossibility and necessity can be represented and reasoned about: they are not extremes of probability!

A recent extension to the M-M project is based on analysis of differences in requirements for logical and algebraic theorem proving, which are now very well established technologies, and requirements for the kinds of reasoning mechanisms that seem to be essential for the ancient topological and geometrical discoveries. Early ideas about this are presented in discussion of a conjectured multi-layer "Super-Turing membrane machine" able to support ancient mathematical reasoning, with precursor versions in other species and pre-verbal human toddlers. It is not clear yet whether these can be implemented as virtual machines running on digital computers or whether something very different is required, e.g. perhaps chemical computation involving a mixture of discrete and continuous processing. Some preliminary thoughts, based on partial analysis of geometric reasoning examples can be found here: http://www.cs.bham.ac.uk/research/projects/cogaff/misc/super-turing-geom.html.

Statistical/probabilistic learning must be contrasted with discovery of *possibilities, necessities* and *impossibilities* (recognized in Kant's *Critique of Pure Reason* (1781) as central to mathematical knowledge). Both forms of learning can be useful but proposers of mechanisms sometimes ignore complementary mechanisms.

Whether the "diagrammatic" mechanisms can be implemented as virtual machines running on digital computing machinery is not yet clear. I suspect Turing's interest in "The chemical basis of morphogenesis" (published in 1952, two years before his death) may have been related to these questions.

Researchers interested in how animals or machines interact with their environments tend to make assumptions that I shall challenge. E.g. it is often assumed that the information used is largely numerical, and that knowledge about the environment is either about particular measurements or about statistical regularities derived from percepts at different times and places.

Intelligent reasoning and decision making can often use the fact that very powerful constraints are connected with certain structures and processes that are *impossible* (A is further than B, B is further than C, and C is further than A) and some connections that are *necessary*, e.g. transitivity of many comparisons, and transitivity of containment. http://www.cs.bham.ac.uk/research/projects/cogaff/misc/impossible.html.

A particular sub-class is concerned with one-one correlations: the basis of concepts of cardinal and ordinal numbers. Despite spurious evidence suggesting that humans and some animals have innate concepts of cardinality, these abilities seem not to develop in children until they are 5 or 6 years old.

The tutorial will present examples, discussed interactively with the audience, raising questions about the mechanisms required for use of the information, and implications for future research. I'll try to draw attention to largely unnoticed mathematical competences that are not innate, but are nevertheless products of deep features of biological evolution. Some of the themes will be summarised in my short talk during the main conference. I believe the talk by Catherine Legg on "The Epistemology of Mathematical Necessity" will also be directly relevant.

Before the tutorial, I'll make available an expanded version of this document, with many more links and references, including examples (with diagrams!), here: http://www.cs.bham.ac.uk/research/projects/cogaff/misc/diagrams-tutorial.html.

## References

1. McCarthy, J., Hayes, P.: Some philosophical problems from the standpoint of AI. In: Meltzer, B., Michie, D. (eds.) Machine Intelligence 4, pp. 463–502. Edinburgh University Press, Edinburgh (1969). http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html
2. Sloman, A.: Interactions between philosophy and AI: the role of intuition and non-logical reasoning in intelligence. In: Proceedings of 2nd IJCAI, pp. 209–226. William Kaufmann, London (1971). Reprinted in Artificial Intelligence, vol. 2, no. 3–4, pp. 209–225 (1971)

3. Schmidhuber, J.: New millennium AI and the convergence of history: update of 2012. In: Eden, A.H., Moor, J.H., Soraker, J.H., Steinhart, E. (eds.) Singularity Hypotheses. The Frontiers Collection, pp. 61–82. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32560-1_4
4. Chappell, J., Sloman, A.: Natural and artificial meta-configured altricial information-processing systems. Int. J. Unconv. Comput. **3**(3), 211–239 (2007)

# Using Verbal Protocols to Support Diagram Design

Thora Tenbrink(✉)

Bangor University, Bangor, Wales, UK
`t.tenbrink@bangor.ac.uk`

**Abstract.** How do we know what people perceive in a diagram? A diagram can be an excellent medium for communicating complex facts and relationships. Users may be able to learn a lot just from a quick glance at a well-designed diagram. Unfortunately, what users take from a diagram may not always be the same as what its designers intended to communicate. This tutorial explores the use of verbal protocol analysis in the area of diagram interpretation, and offer practical support for systematic analysis procedures. This includes a close look at the way people formulate their thoughts about a design, which can reveal underlying conceptualisations and perspectives that the speakers may not be aware of.

**Keywords:** Diagram design · Verbal protocols · Miscommunication
Cognitive Discourse Analysis

## 1 Overview

How do we know what people perceive in a diagram? A diagram can be an excellent medium for communicating complex facts and relationships. Users may be able to learn a lot just from a quick glance at a well-designed diagram. Unfortunately, what users take from a diagram may not always be the same as what its designers intended to communicate. This can have enormous consequences, ranging from misinterpretation of research outputs to false representation in the media, to the point of misguided policy decisions coming from miscommunication of central research insights.

In this tutorial, we will look at the use of verbal protocols as a tool in the diagram design process. The way people talk about a diagram can reveal how they understand it, what they misinterpret, and what kinds of design features could be amended to enhance clarity, ensuring successful communication. This is supported by the methodology of Cognitive Discourse Analysis (CODA; Tenbrink 2015), which uses an in-depth linguistic approach to protocol analysis. Besides the (often quite revealing) content of *what* people say, the features of their language (*how* they say it) point to underlying conceptualisations and aspects that the speakers themselves may not be aware of: their focus of attention, aspects that are taken for granted or perceived as new, levels of granularity or detail, conceptual perspectives and switches between them, inferences and (possibly premature) conclusions, and so on.

The tutorial will start by briefly looking at the kinds of problems that frequently arise in diagram interpretation, such as cognitive biases, misinterpretation of accidental

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 20–23, 2018.
https://doi.org/10.1007/978-3-319-91376-6_4

features (e.g. color, shape, or size of a symbol), and effects of lack of expertise. Then we will turn to the practical aspects of verbal data collection procedures, techniques for data preparation towards systematic analysis, features of content analysis, linguistic feature annotation (specific to CODA), reliability of annotation procedures, and identification of patterns in the results.

## 2 Background

Like other visualisation tools, diagrams can be excellent for facilitating the understanding of complex phenomena, in some areas more so than extensive descriptions in words (Larkin and Simon 1987). Since such tools represent relevant aspects of a situation in a schematic way, they are designed to support the perceiver in drawing strategical conclusions and making decisions based on useful heuristics, without necessarily considering every single aspect of the represented facts (Todd and Gigerenzer 2000).

When looking at a diagram, humans do not perceive all features and elements equally or objectively (as a computer might do). Instead, the perceiver's attention is drawn towards visually salient elements (Fine and Minnery 2009) just as well as towards elements that are pertinent to a current task (Henderson et al. 2009). As a consequence, some aspects may remain entirely outside the perceiver's consciousness.

Understanding these principles is vital for designing the visualisation of information in diagrams in a cognitively supportive way (Fabrikant and Goldsberry 2005). The mere inclusion of relevant information is not sufficient if it is not cognitively accessible in the way needed by the perceiver. Apart from failing to identify vital information, perceivers may also misinterpret the representation – for instance, they might confuse accidental features of a diagram as representing actual states or relationships in the real world. In this respect, conventions and expertise play a major role. Regular users of a particular type of diagram are less likely to be misguided than first-time observers. Likewise, the ability to extract relevant information from expert domain visualisations strongly depends on experience (Jee et al. 2009). Thus, perceivers of displayed information are biased by their background as well as by their perception of relevance in a given context. This will affect the inferences and decisions made on the basis of visualised information, leading to either desired or undesired outcomes.

Unfortunately, phenomena of this kind are not directly accessible to observation, since they concern structures and processes in the mind. Instead, access to internal processes is only possible indirectly, through analysis of external representations and measures. Methods to investigate the principles of understanding visualisations and diagrams along with their effectiveness for users encompass eye movement (Fabrikant and Goldsberry 2005) and sketch map (Jee et al. 2009) analysis, field usability studies (Sarjakoski and Nivala 2005), descriptions and arrow use (Heiser and Tversky 2006), etc.

One readily available external representation of cognition is language. Humans are used to speaking about their thoughts, and can normally express their understanding of a diagram in words. Nevertheless, interviewing people concerning their thoughts, as such, may not be sufficient; associated problems include issues around reliability, validity, and systematicity. Often, information gathered this way does not make its way into scientific reports, even if it serves to inform researchers on an informal basis.

Remedying these issues to some extent, Ericsson and Simon (1993) developed a systematic method to address higher-level cognitive processes by eliciting and analysing verbal protocols produced along with cognitively complex tasks, such as problem solving or decision-making. Think-aloud protocols and retrospective reports provide procedural information that systematically complements other data, such as decision outcomes and behavioural performance results. However, as the analysis of verbal reports in this paradigm typically remains on the content level, in many cases the insights gained in this way still remain illustrative and anecdotal, rather than being treated as substantial evidence. Also, since Ericsson and Simon (1993) primarily aimed to identify the sequence of cognitive steps taken to solve a problem or reach a decision, their approach may not be directly applicable to addressing the interpretation of diagrams. In this domain, the cognitively complex challenges do not necessarily emerge in distinct cognitive steps, even if there is a problem to solve and decisions to be made.

Cognitive Discourse Analysis (CODA; Tenbrink 2015) extends Ericsson and Simon's (1993) approach in several respects. It provides an operationalised way of capturing verbalised content using linguistic insights, particularly from two areas that have been extensively studied and tested in English and some other Indo-European languages: Systemic Functional Linguistics (SFL; Halliday and Matthiessen 2014), and Cognitive Linguistics (CL; Talmy 2000, 2007; Evans and Green 2006). In particular, lexicogrammatical structures in language are systematically related to cognitive structures and processes. This structural fact carries over to principles of language in use: the way we think is related to the way we talk. This is true both generally in terms of what we can do with language, and specifically with respect to what we actually do – for example when verbalising thought related to visualisations such as expert domain diagrams.

When asked to verbalise their thoughts, speakers draw in systematic ways from their linguistic repertory to express currently relevant aspects. Their choices in relation to a cognitively demanding situation reveal crucial elements of their underlying conceptualisations and thought patterns. For instance, seemingly synonymous expressions such as *over* and *above* carry different implications and underlying concepts (Talmy 2007). While *above* clearly refers to the vertical dimension in *There is a poster above the hole in the wall, over* is actually polysemous. In *There is a poster over the hole in the wall* it seems reasonable to infer a functional sense of *covering* – a fundamentally different concept than verticality, and therefore a significant choice in the verbalisation.

In a verbal protocol describing the interpretation of a diagram, subtle differences such as that between *over* and *above* can become critical. It might matter to a high extent whether a particular feature of the diagram is perceived as geometrically *vertical* relative to another, or rather as functional in some sense, such as *covering* it.

A systematic analysis of such linguistic details provides a useful pathway to access cognition, drawing on knowledge about relevant features of language supported by grammatical theory, cognitive linguistic semantics, and other linguistic findings. Although linguistic expertise thus provides useful background, the general approach, to start with, is simple enough to be adopted by non-linguistic experts, with the most important feature being operationalisation and systematisation of language analysis.

The methodological steps of CODA (Tenbrink 2015) are straightforwardly accessible to researchers across disciplines; they will be discussed in detail in this tutorial.

Besides building on established insights about the significance of particular linguistic choices, validating evidence for the relationship between patterns of language use and the associated cognitive processes can be gained by triangulation, i.e., the combination of linguistic analysis with other types of evidence such as memory or behavioural performance data, reaction times, eye movements, decision outcomes, or any other relevant data that can be collected in relation to diagrams. The outcome of a CODA-based analysis combined with such data is then a validated account of systematic features of diagram interpretation that may feed directly into design improvement.

## References

Ericsson, K.A., Simon, H.A.: Protocol Analysis: Verbal Reports as Data. MIT Press, Cambridge (1993)

Evans, V., Green, M.: Cognitive Linguistics. An Introduction. Edinburgh UP, Edinburgh (2006)

Fabrikant, S.I., Goldsberry, K.: Thematic relevance and perceptual salience of dynamic geovisualization displays. In: Proceedings 22nd International Cartographic Conference, A Coruna, Spain, 9–16 July 2005 (2005)

Fine, M.S., Minnery, B.S.: Visual salience affects performance in a working memory task. J. Neurosci. **29**(25), 8016–8021 (2009)

Halliday, M.A.K., Matthiessen, C.M.I.M.: Halliday's Introduction to Functional Grammar, 4th edn. Routledge, London (2014)

Heiser, J., Tversky, B.: Arrows in comprehending and producing mechanical diagrams. Cognit. Sci. **30**, 581–592 (2006)

Henderson, J.M., Malcolm, G.L., Schandl, C.: Searching in the dark: cognitive relevance drives attention in real-world scenes. Psychon. Bull. Rev. **16**(5), 850–856 (2009)

Jee, B.D., Gentner, D., Forbus, K., Sageman, B., Uttal, D.H.: Drawing on experience. In: Taatgen, N.A., van Rijn, H. (eds.) Cognitive Science 2009, Amsterdam, pp. 2499–2504 (2009)

Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. Cognit. Sci. **11**, 65–99 (1987)

Sarjakoski, L., Nivala, A.M.: Adaptation to context—a way to improve the usability of mobile Maps. In: Meng, L., Zipf, A., Reichenbacher, T. (eds.) Map-based Mobile Services, Theories, Methods and Implementations, pp. 107–123. Springer, Berlin (2005). https://doi.org/10.1007/3-540-26982-7_8

Talmy, L.: Toward a Cognitive Semantics, vol. 2. MIT Press, Cambridge (2000)

Talmy, L.: Attention phenomena. In: Geeraerts, D., Cuyckens, H. (eds.) Handbook of Cognitive Linguistics, pp. 264–293. Oxford University Press, Oxford (2007)

Tenbrink, T.: Cognitive discourse analysis: accessing cognitive representations and processes through language data. Lang. Cognit. **7**(1), 98–137 (2015)

Todd, P.M., Gigerenzer, G.: Précis of simple heuristics that make us smart. Behav. Brain Sci. **23**, 727–780 (2000)

# Peirce on Diagrammatic Reasoning and Semeiotic

Javier Legris[1(✉)] and Cassiano Terra Rodrigues[2]

[1] Instituto Interdisciplinario de Economía Política de Buenos Aires,
CONICET-Universidad de Buenos Aires,
Av. Córdoba 2122, C1120 AAQ Buenos Aires, Argentina
`jlegris@retina.ar`
[2] Departamento de Filosofia, Pontifícia Universidade Católica de São Paulo,
R. Monte Alegre, 984 - Perdizes, São Paulo, SP 05014-901, Brazil
`ctrodrigues@pucsp.br`

**Abstract.** Charles Sanders Peirce (1839–1914) is one of the "grounding fathers" of mathematical logic, having developed all of the key formal results of modern logic. He did it firstly (from 1860 on) in the algebraic tradition of mathematical logic stemming from Boole, combining it with the logic of relations, explicitly developed by Augustus De Morgan. From this, Peirce obtained a system that included quantifiers—a term he seems to have invented—and relative predicates. Developing his own system of relative terms, Peirce started from Boole's system, trying to apply it to De Morgan's logic of relations. Indeed, Peirce's aim is to include the logic of relations into the calculus of algebra using his own system of algebraic signs. On the one hand, Peirce's algebraic notation will be presented, specially: (a) relative terms as iconic representations of logical relations; (b) Peirce's quantifiers and the passage from a linear notation to a diagrammatic one. On the other hand, Peirce's graphical notation will be presented, specially: (a) his Alpha and Beta systems, which are fully compatible with what is nowadays called first-order logic, (b) and his unfinished Gamma system, designed for second-order logic and modal logic.

**Keywords:** Relatives · Quantifiers · Diagrammatic Reasoning
Existential Graphs · Semeiotic

## 1 Introduction: Peirce's Early Algebraic Logic

Peirce refuses Boole's identification of logical relations with equations. Holding that inclusion between classes is previous to identity, and *implication* is previous to both, Peirce aims to develop an abstracter calculus so that all logical relations can be defined solely upon the formal characters of one single fundamental relation. So, he replaces the identity sign "=" used by Boole by the specific sign "$-\!\!<$" (the 'craw foot') for the fundamental subsumptive operation of *illation*, which encompasses the logical relations of conditionality, inclusion and consequence [1: 360].

Next, Peirce deals with the composition of relations with classes, and not strictly relational composition, that is, composition of a relation with another, as in De

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 24–27, 2018.
https://doi.org/10.1007/978-3-319-91376-6_5

Morgan's system. So, Peirce works with expressions like "lover of ____", or "giver to ____ of ____", rather than De Morgan's expressions with verbs like "____ loves ____" or "____ gives ____ to ____". Thus, the traditional interpretation of propositions as subject-predicate structures is maintained, but transformed, since not restricted to the predication of only one subject. A proposition is a blank-predicate-form, a kind of *icon*, from which the subject-terms are dropped of, with resulting gaps for the insertion of individual variables, the signs for which are *indexes*. Quantifiers can then be introduced, since to express Boole's algebra in relative terms, particularly hypothetical and particular propositions, existential quantification is needed, e.g., a relative term for *case of the existence of* _____; or for *what exists only if there is not* _____; or else *case of the non-existence of* ____; or still *what exists only if there is not*____ [1: 423].

Peirce's fully-fledged theory of multiple quantification in algebraic notation uses the Greek letters $\prod$—for logical *product*—and $\sum$—for logical *sum*—respectively to designate the *universal quantifier* and the *existential quantifier*. Then, there are juxtaposed subscript letters functioning as indexes, that is, specific deictic signs indicating specific items within the defined universe of discourse. As individual variables, indexes show which terms are bound together by a certain relation and in which specific order. For instance, if $l$ denotes the relation of loving, then $l_{\mathrm{ij}}$ signifies "*i* loves *j*", with "*i"* and "*j"* as indexes for whatever individuals are in this love relation. So, a propositions like, e.g., "Everybody loves Chaplin", can be symbolized as $\prod_i l_{iC}$ (*C* as an index for the individual Chaplin); or else "Everyone loves someone" is rendered as $\prod_i \sum_j l_{ij}$ (*j* for *jemand*, German for *someone*). Peirce elsewhere remarks a proper notation necessarily includes icons and indexes, so $\prod$ and $\sum$ were chosen to make the notation as *iconic* as possible. For him, "every algebraical equation is an icon, in so far as it exhibits, by means of the algebraical signs (which are not themselves icons), the relations of the quantities concerned" [2: 13]. In Peirce's semeiotic, an icon is a sign that formally resembles its object. So, it better conveys the very movement of thought by "carrying the mind from one point to another", e.g., from the premises to the conclusion [2: 10].

Towards the end of the 19th century, Peirce developed a *diagrammatical* system he himself considered his *chef d'oeuvre* in logic: his *Existential Graphs* (EG). Developing it more or less at the same time as his conception of logic as semiotic, that is, the "*quasi*-necessary" and general doctrine of signs, Peirce considered it as *the logic of the future*, abandoning his early algebraic attempts for *philosophical* reasons.

As special kinds of icons, diagrams resemble their objects only in the aspects that attention needs to be drawn upon, that is, "only in respect to the relations of their parts that their likenesses consists" [2: 13]. So, the EG system is more capable than linear notations to lay bare the inferential movement of thought, showing how formal relations can be inter-derived from one another. Generalizing, all logical inferences can be semiotically interpreted as a sort of *diagrammatic experimentation upon signs*, which are essentially iconic. This point inserts Peirce in a long Western tradition of symbolic thought not restricted to linguistic analysis [3].

Peirce's graphical system includes: (a) his Alpha and Beta systems, which are fully compatible with what is nowadays called first-order logic, (b) and his unfinished Gamma system, designed for second-order logic and modal logic. The system has only three rules—scroll, cut, and line of identity—that permit experimentation and transformation of diagrams. The Existential Graphs system is truly a *topovisual* logical system [4], where only the connections and relations between parts are important, the rules of transformation of which make up the laws of the system. From this, Peirce developed inference rules that anticipated more recent and known systems of diagrammatization.

A less known subject is Peirce's distinction between mathematic and logic, which the graphs make more explicit. Mathematics is for Peirce the science that *draws* necessary conclusions from hypothetical diagrammatic structures, while logic is the science of *drawing* necessary conclusions. In other words, mathematics is the most abstract exercise of reasoning itself, based on a principle of parsimony, the most general of all theoretical activities. Logic, in its turn, is a normative science seeking to determine how we ought to reason, with concerns that can be said of a rhetorical nature [5]. Logic analyses reasoning, breaking it in its least constitutive steps to understand its logical movement. So, a formal system of signs has different uses for each science. Notwithstanding the difference, both logic and mathematics find diagrams most profitable, because all necessary reasoning is iconic, as said. As icons exhibiting the logical connexions among relations, diagrams allow for passing from simultaneity to sequentiality. Peirce's distinction between *theorematic* and *corollarial* deductions is understandable in this context: a theorematic deduction consists in adding elements to the diagram to see what would result of such modification. It is thus a creative abductive experimentation upon the diagram, "the heuretic part of mathematical procedure" [6: 49]. In corollarial deduction, the procedure starts from the observation of a diagram such as it is, without any modification, to affirm the conclusion. The conclusion, therefore, is necessarily obtained only without any further adjunction just by logical development of the diagram. Now, for Peirce, "reasoning essentially consists in the observation that where certain relations subsist certain others are found" [7: 164]. The distinction between the two forms of deductive reasoning shows that necessary reasoning is not limited to the strict drawing of consequences, but it is also a constructive activity of formal representations, by means of observing and modifying other such representations. Both mathematics and logic are experimental activities upon signs in general, and diagrams particularly. By studying and experimenting upon diagrams, we come to understand the very *semiotic* nature of mind itself. Thus, Peirce's arguments for iconicity also work for stressing creativity and discovery in mathematical and logical sciences.

## References

1. Peirce, C.S.: Description of a notation for the logic of relatives, resulting from an amplification of the conceptions of Boole's Calculus of logic (1867). In: Moore, E., et al. (eds.) Writings of Charles Sanders Peirce: A Chronological Edition, vol. 2: 1867–1971, pp. 359–429. Indiana University Press, Bloomington (1984)
2. Peirce, C.S.: Of reasoning in general (1895). In: The Peirce Edition Project (ed.) The Essential Peirce: Selected Philosophical Writings, vol. 2: 1893–1913, pp. 11–26. Indiana University Press, Bloomington (1998)

3. Legris, J.: Peirce's diagrammatic logic and the opposition between logic as calculus vs. logic as Universal language. Rev. Port. De Filos. **73**(3/4), 1095–1114 (2017). https://doi.org/10.17990/RPF/2017_73_3_1095
4. Harel, D.: On visual formalisms. In: Glasgow, J., Narayanan, N.H., Chandrasekaran, B. (eds.) Diagrammatic Reasoning: Cognitive and Computational Perspective, pp. 235–271. The AAAI Press/The MIT Press, Menlo Park, Cambridge (1995)
5. Rodrigues, C.T.: The method of scientific discovery in Peirce's philosophy: deduction, induction, and abduction. Log. Univ. **5**(1), 127–164 (2011)
6. Peirce, C.S.: Carnegie application (L 75, 1902). In: Eisele, C. (ed.) The New Elements of Mathematics, vol. 4, pp. 36–73. Mouton Publishers/Humanities Press, The Hague, Atlantic Highlands (1976)
7. Peirce, C.S.: On the algebra of logic: contribution to a philosophy of notation (1885). In: Fisch, M.H., Kloesel, C.J.W., et al. (eds.) Writings of Charles Sanders Peirce: A Chronological Edition, vol. 5: 1884–1886, pp. 161–190. Indiana University Press, Bloomington (1993)

Check for updates

# Picturing Quantum Processes

## A First Course on Quantum Theory and Diagrammatic Reasoning

Bob Coecke[1] and Aleks Kissinger[2](✉)

[1] Quantum Foundations, Logics, and Structures, Department of Computer Science,
Oxford University, Oxford, UK
`coecke@cs.ox.ac.uk`
[2] Quantum Structures and Logic, Institute for Computing and Information Sciences,
Radboud University, Nijmegen, Netherlands
`aleks@cs.ru.nl`
`http://www.cs.ox.ac.uk/people/bob.coecke`
`http://www.cs.ru.nl/A.Kissinger/`

**Abstract.** We provide a self-contained introduction to quantum theory using a unique diagrammatic language. Far from simple visual aids, the diagrams we use are mathematical objects in their own right, which allow us to develop from first principles a completely rigorous treatment of 'textbook' quantum theory. Additionally, the diagrammatic treatment eliminates the need for the typical prerequisites of a standard course on the subject, making it suitable for a multi-disciplinary audience with no prior knowledge in physics or advanced mathematics.

By subscribing to a diagrammatic treatment of quantum theory we place emphasis on quantum processes, rather than individual systems, and study how uniquely quantum features arise as processes compose and interact across time and space. We introduce the notion of a process theory, and from this develop the notions of pure and mixed quantum maps, measurements and classical data, quantum teleportation and cryptography, models of quantum computation, quantum algorithms, and quantum non-locality. The primary mode of calculation in this tutorial is diagram transformations, where simple local identities on diagrams are used to explain and derive the behaviour of many kinds of quantum processes.

This tutorial roughly follows a new textbook published by Cambridge University Press in 2017 with the same title.

## 1 Rationale

The past ten years have witnessed a flurry of activity and change in the way we visualise and reason about quantum processes, coming largely from two directions: the program of *categorical quantum mechanics* initiated by Abramsky and Coecke in 2004, and work in *quantum foundations* and particularly in the *axiomatic reconstruction of quantum theory* a la Hardy and the Pavia group. What these two programs have in common is a shift from studying individual

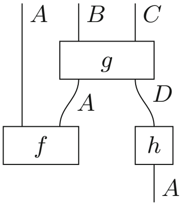
© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 28–31, 2018.
https://doi.org/10.1007/978-3-319-91376-6_6

quantum systems concretely to studying properties of interaction and information flow through compound systems. More concretely, this program has seen a shift from using symbolic languages, where composition and connectivity remain hidden, to diagrammatic ones.

More recently, these category-theory inspired diagrammatic representations have started to be used in many application areas, which include computer science, quantum theory, natural language modelling, and many others. There are now many dedicated conferences and workshops, dedicated research groups, a large international community, and a dedicated journal in the making.

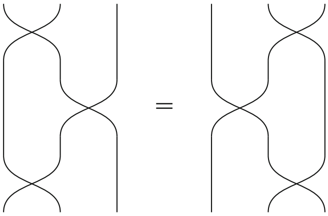
Most of the prior literature required either a working background in the foundations of physics or in category theory, so was not accessible to a wider audience. Our book *Picturing Quantum Processes* is a first exception to this, aiming at a wide multidisciplinary audience. The rational behind this tutorial is to bring several communities together, the one of this conference series, and the ones discussed above, in order to see where there are overlaps and opportunities for cross-fertilisation.

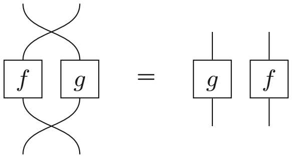
The **language** we use is that of *string diagrams*, which came from mathematical physics as a succinct method of performing calculations in the tensor calculus for (multi-)linear algebra and differential geometry.



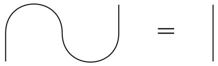
$$\longleftrightarrow \quad f^{A_1A_2}g^{B_1C_1}_{A_2D_1}h^{D_1}_{A_3}$$

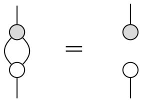
It was later discovered that these form a sound and complete language for describing a very general family of mappings, specifically: morphisms in any monoidal category. Roughly speaking, monoidal categories are the most general definition of a ‘family of processes closed under vertical (i.e. time-like) and horizontal (i.e. space-like) composition’. Thus, string diagrams give an ideal starting point for studying properties of a physical theory while building in as few assumptions as possible. At the same time, a shift to diagrammatic language yields big benefits, by subsuming many non-trivial equations into simple diagram deformations, e.g.



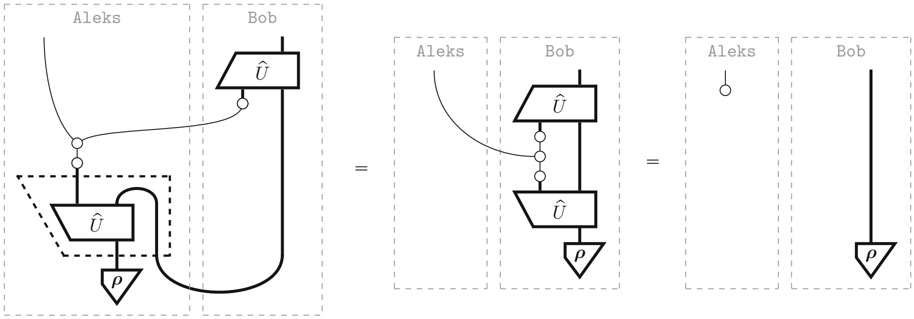


The methodology of the tutorial eschews concrete, low-level calculation wherever possible, choosing instead to highlight how the diagrammatic properties of quantum processes give rise to quantum phenomena. We will see in this tutorial that central concepts such as *non-separability of quantum states* and *complementarity of quantum measurements* will be captured in simple diagram identities:





Using such basic identities, one can express, and prove properties about, more elaborate quantum processes, computations, and communication protocols. For example, the following is a derivation of the quantum teleportation protocol, whereby Aleks communicates an arbitrary quantum state to Bob by means of quantum entanglement (thick wires) and classical communication (thin wires):



This tutorial will introduce the fundamentals of this diagrammatic language, and survey its applications in various areas of quantum theory, computation, and beyond.

No particular pre-requisites are assumed, except a level of general mathematical knowledge. As such, this tutorial is particularly well-suited to the multidisciplinary nature of the participants in the Diagrams conference.

## 2 References

The main reference for this tutorial is the textbook with the same title [1], published in 2017 by Cambridge University Press. A condensed version of approximately the first 2/3 of the book are available in a series of journal articles:

- Categorical Quantum Mechanics I: Causal Quantum Processes. arXiv:1510.05468
- Categorical Quantum Mechanics II: Classical-Quantum Interaction. arXiv:1605.08617

For those interested in some of the categorical underpinnings of this course, the following is some (strictly optional) extra reading: [2–5].

## References

1. Coecke, B., Kissinger, A.: Picturing Quantum Processes. A First Course in Quantum Theory and Diagrammatic Reasoning. Cambridge University Press, Cambridge (2017)
2. Abramsky, S., Coecke, B.: A categorical semantics of quantum protocols. In: Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science (LICS), pp. 415–425 (2004). arXiv:quant-ph/0402130
3. Selinger, P.: A survey of graphical languages for monoidal categories. In: Coecke, B. (ed.) New Structures for Physics. Lecture Notes in Physics, vol. 813, pp. 289–355. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-12821-9_4. arXiv:0908.3347
4. Backens, M.: The ZX-calculus is complete for stabilizer quantum mechanics. New J. Phys. **16**, 093021 (2014). arXiv:1307.7025
5. Coecke, B., Duncan, R.: Interacting quantum observables: categorical algebra and diagrammatics. New J. Phys. **13**, 043016 (2011). arXiv:quant-ph/09064725

# Carroll Diagrams: Design and Manipulation

Amirouche Moktefi(✉)

Tallinn University of Technology, Tallinn, Estonia
amirouche.moktefi@ttu.ee

**Abstract.** The use of diagrams in logic is old. Euler and Venn schemes are among the most popular. Carroll diagrams are less known but are occasionally mentioned in recent literature. The objective of this tutorial is to expose the working of Carroll's diagrams and their significance from a triple perspective: historical, mathematical and philosophical. The diagrams are exposed, worked out and compared to Euler-Venn diagrams. These schemes are used to solve the problem of elimination which was widely addressed by early mathematical logicians: finding the conclusion that is to be drawn from any number of propositions given as premises containing any number of terms. For this purpose, they designed symbolic, visual and sometimes mechanical devices. The significance of Venn and Carroll diagrams is better understood within this historical context. The development of mathematical logic notably created the need for more complex diagrams to represent *n* terms, rather than merely 3 terms (the number demanded by syllogisms). Several methods to construct diagrams for *n* terms, with different strategies, are discussed. Finally, the philosophical significance of Carroll diagrams is discussed in relation to the use of rules to transfer information from a diagram to another. This practice is connected to recent philosophical debates on the role of diagrams in mathematical practices.

**Keywords:** Carroll diagram · Venn diagram · Universe of discourse
Diagram for *n* terms · Rules · Elimination

## 1 Introduction

Logic diagrams have a long history [16]. New schemes are regularly invented to solve old and new logic problems. Euler popularized the scheme which uses circles to represent classes and their relations [6]. Venn, subsequently, introduced a modified version to overcome some difficulties faced by the users of the original Eulerian scheme [17]. Several diagrams were invented in the immediate post-Venn period, among which a very interesting scheme introduced by Carroll [3].

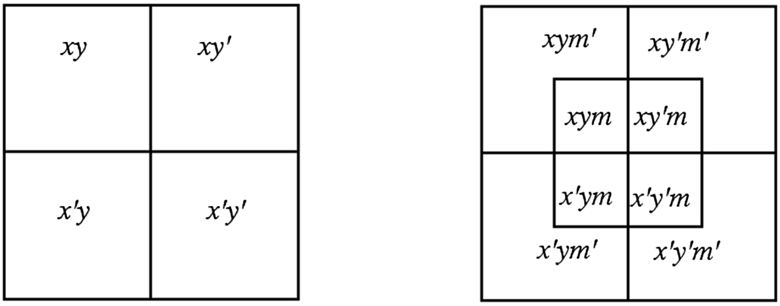
Carroll's diagrams are roughly Venn-type diagrams where the universe is represented with a square. It is not clear whether Carroll worked his scheme as a modification of Venn's. Yet, it stands as a "mature" method summing up several improvements introduced by his predecessors and contemporaries. It represents the universe of discourse [4], it includes a syntactical device for existential import [15] and it has an easy method to construct diagrams for a high number of terms [9, 11]. Although it is not rare to meet with these diagrams in modern textbooks, they have seldom been used in a systematic way together or instead of Euler-Venn diagrams [7, 8].

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 32–35, 2018.
https://doi.org/10.1007/978-3-319-91376-6_7

The objective of this tutorial is to expose the working of Carroll diagrams and their significance from a triple perspective: historical, mathematical and philosophical. In the following, the main design of the diagrams is presented, then, the organisation of the tutorial is sketched to indicate the main directions that are explored.

## 2 Carroll Diagrams

Carroll first represents the universe of discourse with a square, then successive divisions introduce terms (*x*, *y*, *m*, etc.) and their opposites ($x'$, $y'$, $m'$, etc.). For 2 terms, Carroll obtains 4 compartments as shown in the biliteral diagram (Fig. 1a). For 3 terms, one gets the triliteral diagram (Fig. 1b). In order to represent propositions on these diagrams, one has to add syntactical devices: '0' for emptiness and 'I' for occupation (non-emptiness). For instance, to represent the proposition "No *x* is *y*", one simply puts '0' on the *x y* compartment. Similarly, to represent "Some *x* are *y*", one just puts 'I' on the *x y* compartment.



(a) (b)

**Fig. 1.** Carroll's biliteral and triliteral diagrams.

In order to solve a simple problem such as a syllogism, Carroll first represents the premises on the triliteral diagram, then information is transferred to the biliteral diagrams which is supposed to represent the conclusion. This transfer is made by following two main rules:

1- If the quarter of the triliteral diagram has 'I' in either Cell, then it is occupied. One has to mark the corresponding quarter of the biliteral diagram with "I".

2- If the quarter of the triliteral diagram has two "0"s, one in each cell, then it is empty. One has to mark the corresponding quarter of the biliteral diagram with "0".

Suppose we are offered the premises "No *m* is *y*" and "Some *x* are *m*". They are represented on the triliteral diagram in (Fig. 2a). Then, information is transferred to the biliteral diagram (Fig. 2b) which gives the conclusion "Some *x* are not-*y*".

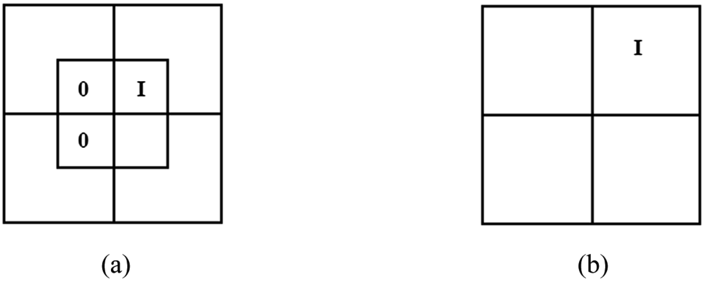


**Fig. 2.** The solution of a syllogism with Carroll diagrams.

## 3 Directions

The tutorial opens with a brief historical introduction to Carroll diagrams, both within the history of logic diagrams and within Carroll's logical theory [1, 2, 10, 12]. The diagrams are then exposed, worked out to solve specific logic problems and compared to Euler-Venn diagrams. This discussion engages in three main directions: historical, mathematical and philosophical.

First, Carroll diagrams are confronted to the problem of elimination which was widely addressed by the mathematical logicians of the time. Indeed, in George Boole's footsteps, logicians worked on general methods for finding the conclusion that is to be drawn from any number of propositions given as premises containing any number of terms. For this purpose, they designed symbolic, visual and sometimes mechanical devices. The significance of Venn and Carroll diagrams is better understood within this historical context.

Second, Carroll diagrams are discussed in relation to the problem of constructing diagrams for *n* terms. Unlike syllogisms which demand only 3 terms, and therefore 3-term diagrams, elimination requires complex diagrams for more than 3 terms. Several methods to construct diagrams for *n* terms, with different strategies were invented in Venn's time [14]. The problem of constructing 'nice' Venn diagrams for *n* terms remained open however and has been recently addressed in mathematical literature [5].

Finally, the philosophical significance of Carroll diagrams is discussed. Venn extracted conclusions 'directly' while Carroll introduced transfer rules. Hence, Venn appealed to imagination to work out the conclusion with a single diagram while Carroll applied rules on a diagram to derive others. The former method can be said to lack rigor, but the latter can be accused of lacking naturalness and economy. This difference of practices, and the philosophical views that they embody, resurfaces in recent debates on the role of diagrams in mathematical practices [13].

## 4 Organization and Practicalities

The tutorial is designed to run on 90 min and is divided in 3 parts. The first part (30 min) provides the historical background to understand the needs and constraints that shaped Euler, Venn and Carroll diagrams. The second part (30 min) exposes Carroll's diagrams, the justification of their design, and their use to work out the elimination problem. Finally, the third part (30 min) discusses the significance of the diagrams and what they teach us on the role of diagrams in mathematical practice.

The tutorial requires only a minimal preliminary knowledge on logic diagrams and their working to solve basic logic problems such as syllogisms. Participants are invited to bring a notebook and pencils to participate to the activities proposed in the tutorial. They will be asked to work out some exercises in logic using Carroll diagrams. No preliminary readings are required but participants are encouraged to consult Carroll's little book *The Game of logic* (1887, several modern reprints and translations).

## References

1. Abeles, F.F.: Lewis Carroll's visual logic. Hist. Philos. Log. **28**(1), 1–17 (2007)
2. Bartley III, W.W. (ed.): Lewis Carroll's Symbolic Logic. C. N. Potter, New York (1986)
3. Carroll, L.: The Game of Logic. Macmillan, London (1887)
4. Coumet, E.: The game of logic: a game of universes. In: Guiliano, E. (ed.) Lewis Carroll Observed, pp. 181–195. C. N. Potter, New York (1976)
5. Edwards, A.W.F.: Cogwheels of the Mind: The Story of Venn Diagrams. Johns Hopkins University Press, Baltimore (2004)
6. Euler, L.: Lettres à une princesse d'Allemagne, vol. 2. Imprimerie de l'Académie impériale des sciences, Saint Petersburg (1768)
7. Geach, P.: Reason and Argument. University of California Press, Berkeley (1976)
8. Lewis, C.I.: A Survey of Symbolic Logic. University of California press, Berkeley (1918)
9. Macula, A.J.: Lewis Carroll and the enumeration of minimal covers. Math. Mag. **68**(4), 269–274 (1995)
10. Moktefi, A.: Lewis Carroll's logic. In: Gabbay, D.M., Woods, J. (eds.) British Logic in the Nineteenth Century, pp. 457–505. North-Holland, Amsterdam (2008)
11. Moktefi, A.: Beyond syllogisms: Carroll's (marked) quadriliteral diagram. In: Moktefi, A., Shin, S.-J. (eds.) Visual Reasoning with Diagrams, pp. 55–71. Basel, Birkhäuser (2013)
12. Moktefi, A.: On the social utility of symbolic logic: Lewis Carroll against 'The Logicians'. Stud. Metodol. **35**, 133–150 (2015)
13. Moktefi, A.: Diagrams as scientific instruments. In: Benedek, A., Veszelszki, A. (eds.) Virtual Reality – Real Visuality, pp. 81–89. Peter Lang, Frankfurt am Main (2017)
14. Moktefi, A., Bellucci, F., Pietarinen, A.-V.: Continuity, connectivity and regularity in spatial diagrams for N terms. In: Burton, J., Choudhury, L. (eds.) DLAC 2013: Diagrams, Logic and Cognition, CEUR Workshop Proceedings, vol. 1132, pp. 31–35 (2014)
15. Moktefi, A., Pietarinen, A.-V.: On the diagrammatic representation of existential statements with Venn diagrams. J. Logic Lang. Inf. **24**(4), 361–374 (2015)
16. Moktefi, A., Shin, S.-J.: A history of logic diagrams. In: Gabbay, D.M., Pelletier, F.J., Woods, J. (eds.) Logic: A History of Its Central Concepts, pp. 611–682. North-Holland, Amsterdam (2012)
17. Venn, J.: On the diagrammatic and mechanical representation of propositions and reasonings. Philos. Mag. **10**, 1–18 (1880)

# Generating and Drawing Euler Diagrams

# Generating Effective Euler Diagrams

Almas Baimagambetov[1(✉)], John Howse[1], Gem Stapleton[1], and Aidan Delaney[1,2]

[1] Centre for Secure, Intelligent and Usable Systems,
University of Brighton, Brighton, UK
{a.baimagambetov,john.howse,g.e.stapleton}@brighton.ac.uk
[2] University of the South Pacific, Suva, Fiji
aidan@ontologyengineering.org

**Abstract.** Euler diagrams are used for visualizing categorized data, with applications including crime control, bioinformatics, classification systems and education. Various properties of Euler diagrams have been empirically shown to aid, or hinder, their comprehension by users. Therefore, a key goal is to automatically generate Euler diagrams that possess beneficial layout features whilst avoiding those that are a hindrance. The automated layout techniques that currently exist sometimes produce diagrams with undesirable features. In this paper we present a novel approach, called iCurves, for generating Euler diagrams alongside a prototype implementation. We evaluate iCurves against existing techniques based on the aforementioned layout properties. This evaluation suggests that, particularly when the number of zones is high, iCurves can outperform other automated techniques in terms of effectiveness for users, as indicated by the layout properties of the produced Euler diagrams.

## 1 Introduction

Visualizing data can be highly effective due to the human brain processing visual information significantly quicker than textual information [19]. Data items that are grouped into sets can be visualized by a variety of techniques, including LineSets [1] and linear diagrams [9,18], with by far the most prominent being Euler diagrams [2]. They are widely used for visualizing sets in numerous application areas, including genetics [11], crime control [6], education [10] and classification systems [24]. The Euler diagram in Fig. 1 conveys information about modules being studied by students. For instance, those who study Web also study Architecture and those who study Logic all study Maths but not Architecture or Programming. This diagram has features that are known to be effective for cognition, such as circular curves each of which is a unique colour [3]. Empirical studies have identified (in)effective features (further details are given later) including their so-called well-formedness properties [16], that are topological in nature, as well as their graphical properties [3]. In the context of automated layout tools, the difficulty of producing visualizations is compounded by the desire to produce effective drawings.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 39–54, 2018.
https://doi.org/10.1007/978-3-319-91376-6_8

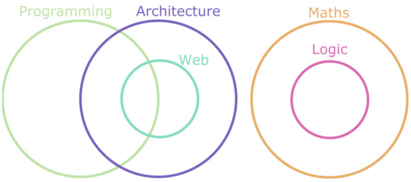


**Fig. 1.** An Euler diagram.

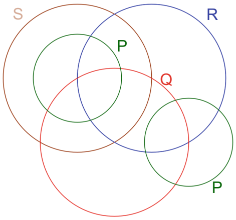


**Fig. 2.** Venn-4: iCircles.

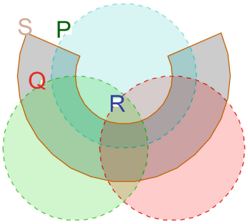


**Fig. 3.** Venn-4: GED-L.

Given the extensive practical uses of Euler diagrams, it is unsurprising that many automated layout techniques exist for drawing them, with the first one being produced in 2002 [7]. However, all existing methods have limitations, such as not being able to represent all data sets, producing diagrams that inaccurately represent the data, or breaking well-formedness properties. Indeed, it is impossible to exactly visualize some sets in a completely well-formed way [7]. Consequently, layout methods have to prioritise which desirable features to enforce in their layouts and which to compromise. It is unknown which Euler diagram layout method produces the most effective diagrams.

The contribution of this paper is two-fold. Firstly, we introduce a new general Euler diagram drawing technique, called iCurves, that aims to ensure desirable properties hold (the diagrams are well-formed) at the expense of sometimes introducing extra zones into the diagrams. Secondly, we empirically compare the diagrams produced by iCurves against selected state-of-the-art techniques to gain insight into their relative effectiveness. Section 2 gives a brief summary of effective layout properties and existing techniques for automated Euler diagram layout. In Sect. 3 we cover the concrete (drawn) and abstract Euler diagram syntax. Section 4 presents the theoretical underpinning of our method, iCurves, which we have implemented. A comparative evaluation of iCurves with existing techniques, using real-world data, is given in Sect. 5. We conclude and discuss future work in Sect. 6. The iCurves software and the diagrams on which our evaluation is conducted can be found at https://github.com/AlmasB/d2018.

## 2 Euler Diagrams: Background

As indicated in the introduction, effective Euler diagrams should be *well-formed* and possess other desirable graphical properties such as the use of circles and colours that are perceptually distinguishable. Five *well-formedness properties* have been identified and established to impact user understanding [16], which are as follows. *Simple curves:* no curve self-intersects. *Non-concurrent curves:* no parts of the curves run along the same path. *Only two-points:* whenever a point is passed through by curves, it is passed through at most twice; points that fail this condition are called *triple points*. *Connected zones:* all of the zones[1]

---

[1] A zone is a maximal region of the plane inside a subset of the curves and outside the remaining curves.

in the diagram are connected components of the plane; a zone which is not connected is called *disconnected*. *Distinct labels:* no two curves have the same label; curve labels that occur more than once are called *duplicated curve labels*. An example with duplicated curve labels is in Fig. 2, which depicts Venn-4 (a Venn diagram representing four sets) using two curves for the set $P$; it was drawn using iCircles [21], with labels manually added for readability. A diagram which does not break any well-formedness property is called *well-formed*.

Rodgers et al. performed two comparative user studies which revealed that Euler diagrams with duplicated curve labels gave rise to significantly worse performance than those without [16]. They also found that disconnected zones caused significantly worse performance than the remaining properties. Thus, it is particularly important that drawing algorithms avoid disconnected zones and duplicated curve labels. The remaining well-formedness properties were shown to cause significantly worse performance using diagrams as compared to those which are well-formed. So, they should also be avoided where possible, but not at the expense of disconnected zones or duplicated curve labels. Therefore we can see, from the perspective of well-formedness, that Fig. 2 is less desirable than Fig. 3, which was generated using the General Euler Diagram drawing method in [15] extended to include a library of simple cases [17]; we call the extended method GED-L.

It has been suggested that the use of curves of particular geometric shapes impacts effectiveness [3]: circles were found to be more effective than ellipses, squares or rectangles. The choice of shape is a core part of layout algorithms, as well as routing the curves in order to display the correct set of zones. This is certainly no easy task and the resulting efficacy of the drawn diagrams is inherently intertwined with the design of the algorithms. By contrast, the use of colour applied to a diagram's curves is easily altered post-layout. Therefore, in this paper we will judge the efficacy of diagram layouts, in Sect. 5, by focusing on the five well-formedness properties that impact understanding and the use of circles. We acknowledge that this does not take into consideration diagram aesthetics but there are currently no comprehensive results that indicate users' aesthetic preferences. One of our goals is to provide insight into the relative effectiveness of the diagrams produced by automated layout tools using known results concerning task performance, not user preference.

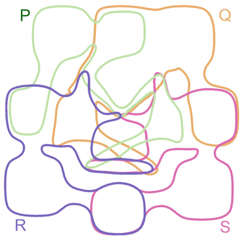


**Fig. 4.** Venn-4: Bubble Sets.

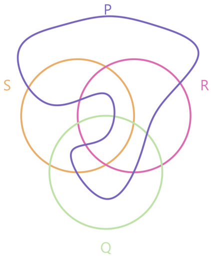


**Fig. 5.** Venn-4: iCurves.

As well as iCircles and GED-L, a large number of other layout tools for Euler diagrams have been provided, such as [5,14,20,23]; see [2] for a comprehensive overview. All of these techniques have problems. For example, the first Euler diagram drawing method only produces well-formed diagrams but many sets cannot be visualised in this way [7]. Moreover, Bubble Sets [5] produces concurrency and disconnected zones (see Fig. 4). Ultimately, there is no ‘ideal’ technique and it is necessary to accept some ineffective features in Euler diagram layouts. The ambition should be to minimise the ineffective features, whilst ensuring the diagram represents the desired sets. Our technique, iCurves (see Fig. 5), sets out to avoid layout features known to be detrimental to task performance.

## 3 Euler Diagrams

We present prerequisite theory for iCurves. The definitions are the same as, or adaptations of, the material, found in [21–23]. An Euler diagram is a collection of closed curves, each of which has a label from some given set of labels, $\mathcal{L}$.

**Definition 1.** *An* ***Euler diagram*** *is a pair,* $d = (C, l)$*, where* $C$ *is a finite set of closed curves drawn in the plane, each curve has a non-empty interior, and* $l\colon C \to \mathcal{L}$ *is an injective function that returns the label of each curve.*

Note that this definition is, for simplicity, more strict than some given for Euler diagrams, such as [21], which do not require non-empty curve interiors or $l$ to be injective (i.e. when duplicated curve labels are allowed).

The closed curves partition the plane into *minimal regions*, which are connected components of the plane. In Fig. 6, the eight minimal regions are enumerated. Another important concept is that of *zones*. A zone is a set of minimal regions that is inside certain curves (possibly none) and outside the remaining curves [23]. In Fig. 6, minimal regions 7 and 8 form a single (disconnected) zone, inside $P$ and $Q$, but outside $R$, representing the set $(P \cap Q)\backslash R$. In total, there are seven zones. A minimal region is a purely topological notion related to a drawn Euler diagram, whereas zones represent the intersection of sets.

Euler diagram drawing techniques, including iCurves, produce diagrams from *descriptions*. A description includes a list of *abstract zones*, where each abstract zone describes a zone. For example, in Fig. 6, the zone formed from minimal regions 7 and 8 is described by $\{P, Q\}$ since it is inside $P$ and $Q$ only.

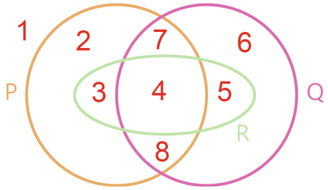


**Fig. 6.** Minimal regions and zones.

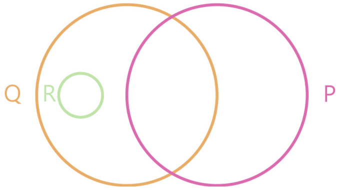


**Fig. 7.** A nested Euler diagram.

**Definition 2.** *A **description**, $D$, is a pair, $(Z, L)$, where $L$ is a subset of $\mathcal{L}$ and $Z \subseteq \mathbb{P}(L)$ such that $\emptyset \in Z$ and each label in $L$ appears in at least one **abstract zone** in $Z$. We define $Z(D) = Z$ and $L(D) = L$.*

In Fig. 7, the diagram, $d$, has a zone inside the curve $P$ only, described by the abstract zone $\{P\}$, and a zone inside precisely $P$ and $Q$, described by $\{P, Q\}$ and so on. So the description, $D$, of $d$ has labels $L = \{P, Q, R\}$ and abstract zones $Z = \{\emptyset, \{P\}, \{Q\}, \{P, Q\}, \{Q, R\}\}$, where $\emptyset$ describes the zone outside all curves. We say that $d$ is a *drawing* of $D$. Furthermore, we will abuse notation for conciseness and write $D$ as $\{\emptyset, P, Q, PQ, QR\}$.

Some automated layout techniques, such as iCircles, draw diagrams with *additional* zones which are not indicated in the description. Such extra zones can be shaded, following the use of shading in Venn diagrams [25], to show that they represent empty sets. Our technique, iCurves, includes extra zones and uses shading in this way.

The concept of *nesting* is of use to us [8]. If the curves of an Euler diagram, $d$, form more than one connected component of the plane, as in Fig. 7 where the curve $R$ is separate from $P$ and $Q$, then $d$ is *nested*, otherwise $d$ is *atomic* (as in Fig. 6). Each connected component is called an *atomic component* [22]. So Fig. 7 has two atomic components, whereas Fig. 6 has one. Given a description, $D$, it is possible to identify *abstract atomic components* [22]. These abstract components correspond to atomic components in some drawing of $D$. In Fig. 7, the two atomic components have the following (atomic) descriptions: $\{\emptyset, P, Q, PQ\}$ and $\{\emptyset, R\}$. Whilst this identification is an essential part of our drawing algorithm, we omit the (complex) theoretical details. What is important here is that the abstract atomic components can be computed. Our algorithm draws the atomic components separately and combines them to form the final diagram. We will refer to atomic and nested descriptions in the obvious way.

## 4 The iCurves Euler Diagram Generation Technique

The goal of the iCurves method is to draw well-formed diagrams, using circles where possible and with extra zones if needed. The drawing process of an abstract atomic component is described by the following sequence of steps:

1. Produce a *decomposition* of the abstract atomic component.
2. Draw the first curve as a circle.
3. Draw the next curve either as a circle or using a *modified dual graph* of the diagram so far. Repeat this step until all curves are drawn.

Given a description, $D$, a *decomposition* of $D$ is a sequence of descriptions, $(D_n, ..., D_0)$, each with one fewer curve label than the previous, where $D_n = D$ and $L(D_0) = \emptyset$. In addition, $D_i = (Z_i, L_i)$ is formed from $D_{i+1} = (Z_{i+1}, L_{i+1})$ by removing a curve label, $\lambda_{i+1}$, from $L_{i+1}$, (so $L_i = L_{i+1} \setminus \{\lambda_{i+1}\}$) and from each abstract zone in $Z_{i+1}$ to give $Z_i$; we write $D_i = D_{i+1} - \lambda_{i+1}$. One of our goals is to sensibly choose a decomposition that allows us to build an effective diagram. A reversed decomposition is called a *recomposition* and corresponds to the order of curve addition described in step 3 above. The use of a decomposition in Euler diagram generation is not new [21,23], but the order of curve label removal can have a profound impact on the effectiveness of the final diagram, as can the method used to draw the curves. We introduce a new strategy for producing a decomposition that takes into account the process by which we draw curves. We also introduce a new technique for drawing the curves themselves.

### 4.1 Drawing Curves in Euler Diagrams

Here, we demonstrate how to draw a new non-circular curve given an existing drawing of an atomic Euler diagram, $d$. This process constructs a *modified dual graph* [23] of $d$ and then finds an *appropriate* cycle for curve addition. To illustrate, given Fig. 8, we produce an *Euler graph* [4], shown in Fig. 9. From this a dual graph is created, shown in Fig. 10. The dual is modified to yield the graph in Fig. 11. The cycle highlighted in Fig. 11 can be used to add a new curve, $S$, as shown in Fig. 12.

*Constructing a Modified Dual Graph.* First, we obtain an *Euler graph*, $EG(d)$, from an Euler diagram, $d$: vertices are points at which the curves cross and the edges are the curve segments that connect the vertices. Next, we obtain a (standard) dual, $EGD(d)$, of an Euler graph. Finally, from $EGD(d)$ we obtain a *modified Euler dual*, $MED(d)$: for each edge, $e$, incident with the vertex, $v_\emptyset$, in the zone outside all of the curves, a new vertex is placed onto $e$; $v_\emptyset$ is deleted along with its incident edges; new edges are added which join the newly inserted vertices, so that the new vertices together with these new edges form a simple cycle that properly encloses the Euler graph.

*Using a Cycle for Curve Addition.* Cycles in $MED(d)$ are used to add curves. In essence, a cycle's edges become curve segments which combine to form a curve. Our method is to use a *simple cycle*, $s$, to add new curves: a simple cycle does not pass through any vertex more than once.

**Definition 3 (adapted from [23]).** *Let $d = (C, l)$ be an atomic Euler diagram with a modified Euler dual $MED(d)$. Let $s$ be a simple cycle in $MED(d)$ and let $\lambda$ be a label not already used in $d$. Then $d$* ***extended by $s$ and $\lambda$*** *is an Euler diagram, denoted $d + (s, \lambda)$, where $d + (s, \lambda) = (C \cup \{c\}, l \cup \{(c, \lambda)\})$ such that $c$ is a closed curve, not in $C$, that traverses the cycle $s$ and has label $\lambda$.*

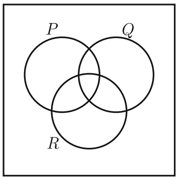


**Fig. 8.** Venn-3.

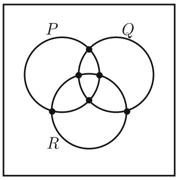


**Fig. 9.** An Euler graph.

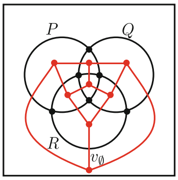


**Fig. 10.** A dual.

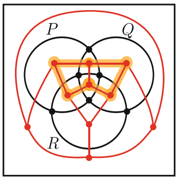


**Fig. 11.** A modified dual.

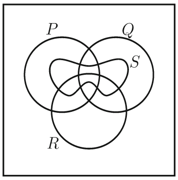


**Fig. 12.** Added $S$.

Of course we cannot use any simple cycle to add a curve, since we have a description that we are aiming to draw. Consider

$$D = \{\emptyset, P, Q, R, PQ, QR, PR, PQR, PS, PQS, QS, QRS, PRS\}$$

which will be used as a running example; it is drawn, with an extra zone, $PQRS$, in Fig. 12. Any decomposition of $D$ is of the form $(D_4, D_3, D_2, D_1, D_0)$, where $D_4 = D$. Suppose that the first label we remove is $S$, giving $D_3 = \{\emptyset, P, Q, R, PQ, QR, PR, PQR\}$. A drawing, $d_3$, of $D_3$ can be seen in Fig. 8, alongside its modified Euler dual, Fig. 11. We now illustrate how we find a cycle in $MED(d_3)$ to create a drawing (with an extra zone) of $D_4$:

1. Identify the abstract zones in $D_4$ that contain $S$: $\{PS, PQS, QS, QRS, PRS\}$. Remove $S$ from these abstract zones to give $\{P, PQ, Q, QR, PR\}$. These abstract zones necessarily occur in $D_3$.
2. Identify vertices in $MED(d_3)$ (Fig. 11) that are placed in the zones with descriptions $\{P, PQ, Q, QR, PR\}$.
3. Next, seek a cycle that passes through at least these vertices. No cycle passes through exactly these vertices, but the cycle, which we call $s$, highlighted in Fig. 11, passes through them all. Use $s$ to add the curve, seen in Fig. 12.

We call the diagram, $d_4$, in Fig. 12, a *relaxed drawing* of $D_4$, as it contains an extra zone; iCurves will shade this non-required zone.

**Definition 4.** *Let $d$ be an atomic Euler diagram. Let $ZON$ be a non-empty set of zones in $d$. Let $s$ be a simple cycle in $MED(d)$. If (i) vertices that arise from zones in $ZON$ are in $s$ and (ii) all vertices in $s$ that are located in the zone outside all curves form one path in $s$ then $s$ is a* ***curve-adding cycle*** *for $d$* ***respecting*** *$ZON$.*

Importantly, curves added using curve-adding cycles ensure that *the resulting diagram is well-formed*, provided the diagram to which the curve is added is also well-formed. In our running example, the cycle used to create $d_4$ (Fig. 12) respects $ZON$, where $ZON$ contains the five zones with descriptions $\{P, PQ, Q, QR, PR\}$. In Fig. 11, in addition to the cycle we used to add the curve $S$, we observe that there exist other curve-adding cycles for $d_3$ respecting $\{P, PQ, Q, QR, PR\}$. The choice of cycle profoundly impacts the layout of the final diagram. For example, using an alternative cycle to add the new curve could result in the diagram $d'_4$, in Fig. 13, which contains three extra zones. We note that the more vertices included in the chosen cycle, the more extra zones the resulting diagram will have. Therefore, iCurves strategically chooses a cycle respecting $ZON$ which has few vertices. We now formalize the notions of a super description and a relaxed drawing.

**Definition 5.** *Let $D_1 = (Z_1, L)$ and $D_2 = (Z_2, L)$ be descriptions. If $Z_1 \subseteq Z_2$ then $D_2$ is a* ***super description*** *of $D_1$.*

**Definition 6.** *Let $D_1$ be a description with super description $D_2$. If $d_2$ is a drawing of $D_2$ then $d_2$ is a* ***relaxed drawing*** *of $D_1$.*

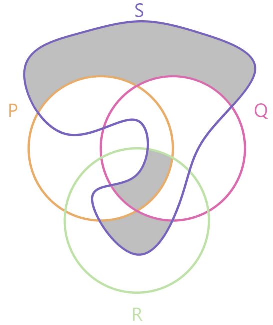


**Fig. 13.** A relaxed drawing.

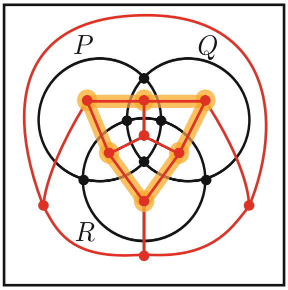


**Fig. 14.** A non-valid cycle.

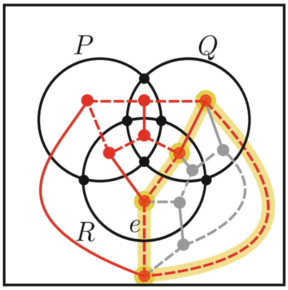


**Fig. 15.** A Hamiltonian dual. (Color figure online)

*H-extendible Euler Diagrams.* It is important that iCurves can draw any description. Yet some curve-adding cycles respecting a given *ZON* set need not ensure we can add the next curve to the resulting diagram. We observe that given a pair of descriptions, $(D_i, D_{i+1})$, in a recomposition, the existence of a Hamiltonian cycle in the Euler graph dual of $d_i$ (a relaxed drawing of $D_i$) ensures we can find a cycle in the modified Euler dual to add a curve that splits each zone in two *and* ensures the resulting Euler graph dual is Hamiltonian. Once a curve is added using a Hamiltonian cycle, we can shade any extra zones to give a relaxed drawing of $D_{i+1}$. Clearly, we want to use our curve adding strategy to minimize the number of extra zones. So the chosen cycle need not be Hamiltonian. However, the presence of a Hamiltonian cycle guarantees that there exists at least one curve-adding cycle and, thus, ensures the drawability of any $D_{i+1}$ given $(D_i, D_{i+1})$. We want to ensure that the Hamiltonian property is preserved.

**Definition 7.** *Let $d$ be an atomic Euler diagram. If an Euler graph dual of $d$ is Hamiltonian then $d$ is* ***H-extendible****.*

In this context, we place an additional constraint on curve-adding cycles. Such cycles partition the plane into two subspaces: bounded and unbounded. A vertex, $v$, is *inside* $s$ if $v$ is in the bounded subspace. It is possible for cycles that have vertices inside them not to give rise to diagrams with Hamiltonian duals. For example, in Fig. 14, if we use the highlighted cycle to add a new curve then the resulting diagram will not have a Hamiltonian dual.

In addition, we place a stronger condition on edges of curve-adding cycles, which guarantees that, after adding a curve, the resulting diagram has a Hamiltonian dual. In Fig. 15, the Euler graph dual includes a Hamiltonian cycle, shown using dashed (red) edges. A curve-adding cycle, $s$, in the Euler graph dual is highlighted by a semi-translucent path (in yellow). If $s$ is used to add a new curve, the Euler graph dual would need to be updated for the new diagram. The resulting new edges and vertices are shown in grey. The dashed grey edges together with the dashed red edges, excluding $e$, form a Hamilitonian cycle in the new Euler graph dual. Notice the edge $e$ and how its incident vertices are used to join the red and grey paths to form the new Hamiltonian cycle thus extending the Hamiltonian property to the new dual. The presences of $e$ in a Hamiltonian cycle is important: it ensures that we can join the dashed paths together using its incident vertices to create a new Hamiltonian cycle. In general, the existence of an edge in both a Hamiltonian cycle in $EGD(d)$ *and* in the curve-adding cycle, is sufficient to ensure that a Hamiltonian cycle exists in the new dual after adding a curve.

**Definition 8.** *Let $d$ be an atomic Euler diagram. Let $ZON$ be a non-empty set of zones in $d$. A curve-adding cycle, $s$, for $d$ respecting $ZON$ is* ***valid*** *if there are no vertices inside $s$ and there is an edge, $e$, in $s$ such that there exists a Hamiltonian cycle in $EGD(d)$ that contains $e$.*

**Lemma 1.** *Let $d$ be an atomic Euler diagram that is $H$-extendible. Let $ZON$ be a non-empty set of zones in $d$. Then there exists a valid curve-adding cycle for $d$ respecting $ZON$.*

It follows that at each step in our diagram generation process we want to generate an $H$-extendible diagram, so that there exists at least one valid curve-adding cycle for the next step.

**Theorem 1.** *Let $(D_i, D_{i+1})$ be a pair of descriptions in a recomposition. Let $d_i = (C, l)$ be an $H$-extendible Euler diagram which is a drawing of $D_i$. There exists an $H$-extendible relaxed drawing, $d_{i+1}$, of $D_{i+1}$ and obtained from $d_i$ by adding a curve using a valid curve-adding cycle.*

We can readily show that given any description, $D$, there exists a relaxed drawing of $D$ by appealing to Venn diagrams, which are known to have Hamiltonian duals [13]. For example, in Fig. 16 we can shade any zone in a Venn-5 diagram and obtain a relaxed drawing of any description with five curve labels.

**Lemma 2.** *Let $D$ be a description. There exists an atomic Euler diagram that is $H$-extendible which is a relaxed drawing of $D$.*

In summary, to guarantee drawability using iCurves, we select the *shortest* valid curve-adding cycles respecting $ZON$ which ensures the resulting diagram is $H$-extendible. Shortest cycles lead to fewer extra zones being introduced when producing a relaxed drawing. The validity and $H$-extendible conditions ensure that we can continue adding curves to create a well-formed relaxed drawing of the required description.

### 4.2 Producing a Decomposition

We have now seen how iCurves adds curves, taking into account extra zones and drawability. The choice of decomposition also impacts on the quality of the final diagram. We present our strategy that leads to a particular choice for the order of curve label removal. Note that given an atomic description, $D$, iCurves draws an atomic diagram. Our strategy ensures that all descriptions in the decomposition of $D$ are atomic. This is achieved by always choosing to remove a label that results in an atomic description, i.e. a *non-disconnecting* label.

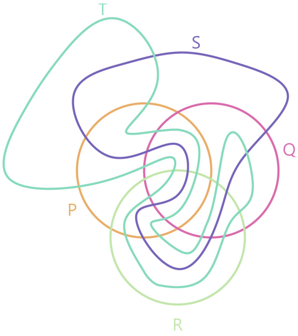


**Fig. 16.** Venn-5.

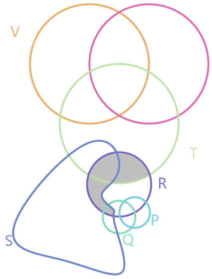


**Fig. 17.** A bad choice of decomposition.

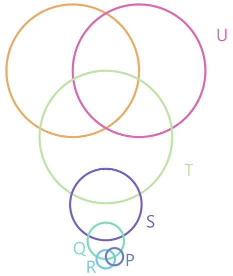


**Fig. 18.** A good choice of decomposition.

**Lemma 3.** *Let $D$ be an atomic description. There exists a label, $\lambda$, in $D$ that, when removed, results in an atomic description.*

This is an important part of our strategy, since removing disconnecting labels, yielding a nested description, necessarily leads to extra zones being included when drawing the diagram. For example, in Figs. 17 and 18, the diagrams were generated from the same description. In Fig. 17 the diagram has four extra zones and was drawn using a decomposition which first removed label $S$, giving a nested diagram; iCurves will not produce this layout as the decomposition violates our strategy. Figure 18 was drawn with iCurves using only atomic descriptions in the decomposition and has no extra zones. Our decomposition strategy to remove a label from $D_{i+1}$ to give $D_i$ is as follows:

1. If there is only one non-disconnecting label in $D_{i+1}$, remove it.
2. Else, see if any non-disconnecting label might be drawable as a circle that 'cuts across' one curve: for each non-disconnecting label, $\lambda$, see if $\{az \in Z(D_{i+1}) : \lambda \in az\} = \{az, az \cup \{\lambda_1\}\}$ for some $az \in Z(D_{i+1})$ and $\lambda_1 \in L(D_{i+1})$. If there are such labels, randomly choose one of them to remove.
3. Else, see if any non-disconnecting label might be drawable as a circle that 'cuts across' two curves, possibly adding one or two extra zones: for each non-disconnecting label, $\lambda$, see if $\{az \in Z(D_{i+1}) : \lambda \in az\} \subseteq \{az, az \cup \{\lambda_1\}, az \cup \{\lambda_2\}, az \cup \{\lambda_1, \lambda_2\}\}$ for some $az \in Z(D_{i+1})$ and $\lambda_1, \lambda_2 \in L(D_{i+1})$. If there are such labels, randomly choose one of them to remove.
4. Else, remove a label whose addition is likely to introduce the fewest extra zones: for each non-disconnecting label, $\lambda$, first compute $n = |\{az \in Z(D_{i+1})\colon \lambda \in az \wedge az \setminus \{\lambda\} \notin Z(D_{i+1})\}|$, which gives the number of abstract zones that contain $\lambda$ and do not have *neighbours*[2]. Next, compute $m$ by finding the smallest set of abstract zones, $Z'$, where $Z' \subseteq Z(D_{i+1} - \lambda)$ such that the graph $(V, E)$, defined by $V = \{az \in Z(D_{i+1} - \lambda)\colon az \cup \{\lambda\} \in Z(D_{i+1})\} \cup Z'$ and $E = \{\{az, az'\}\colon az = az' \cup \{\lambda'\}\}$ for some $\lambda' \in L(D_{i+1} - \lambda)$, is a simple

---

[2] Two abstract zones are neighbours if their symmetric difference has one element.

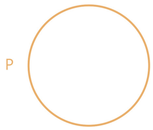


**Fig. 19.** $d_1$.

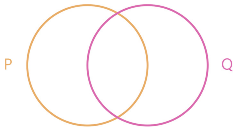


**Fig. 20.** $d_2$.

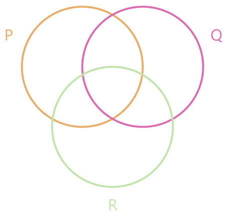


**Fig. 21.** $d_3$.

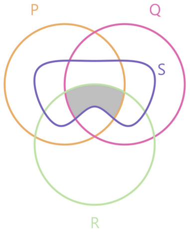


**Fig. 22.** $d_4$.

cycle; set $m = |Z'|$. Remove the label with the smallest $n + m$; if there is a choice, randomly choose one of them to remove.

A decomposition of our running example, $D$, can be formed by removing $S$, then $R$, followed by $Q$ and $P$. The drawings of $D_1$–$D_3$ are in Figs. 19, 20 and 21. The diagram $d_4$, in Fig. 22, is a relaxed drawing of $D_4$ and, therefore, of $D$. In $D$, all four labels are non-disconnecting and cannot be identified as (possibly) drawable with circles. Hence, we compute $n + m$ for each label. The value of $n$ for $P$ is 2 since there are two abstract zones, $PS$ and $PRS$, where corresponding abstract zones, $S$ and $RS$, are not in $D$. To compute $m$ for $P$, we obtain the set $V$, which is formed from two sets $V' \cup Z'$, where $V'$ arises from abstract zones in $D$ that contain $P$. So, we have $V' = \{QR, R, \emptyset, Q, QS, S, RS\}$. The smallest set that can be added to $V'$ to give a simple cycle is $\{QRS\}$. Thus, $m$ for $P$ is $|\{QRS\}| = 1$ and we have $n + m$ for $P$ is 3. Similarly, $n + m$ for $Q$, $R$ and $S$ is 3, 2 and 1 respectively. $S$ has the smallest $n + m$ and, so, is removed from $D$. Next, $P$, $Q$ and $R$ can be removed in any order as they are all drawable as circles.

### 4.3 Drawing Algorithm

To generate an atomic Euler diagram from an atomic description, $D$, using iCurves, call Algorithm 1 (page 12) with $D$ as input.

**Theorem 2.** *Let $D$ be an atomic description. Applying Algorithm 1 to $D$ produces an atomic Euler diagram, $d$, that is a relaxed drawing of $D$.*

To draw a nested description, $D$, iCurves starts by splitting $D$ into its atomic parts, say $D_1$, $D_2$, ..., $D_n$. Each $D_i$ is then drawn using Algorithm 1. The diagrams are then displayed appropriately, reflecting the nested properties of $D$; the formal details are omitted due to space constraints. An example of a nested diagram drawn by iCurves can be seen in Fig. 23. We are now in a position to state our main theorem:

**Theorem 3.** *Let $D$ be a description. Then iCurves produces an Euler diagram, $d$, that is a relaxed drawing of $D$.*

**Algorithm 1.** Draw atomic component.

```
Input  : An atomic description, $D$.
Output: An atomic Euler diagram, $d$, which is a relaxed drawing of $D$.
begin
    Set $dec(D) = (D_n, ..., D_0)$, removing labels using the strategy given above.
    $D_0$ has no labels and $D_1$ has a single label, $\lambda_1$. Draw the first curve as a
    circle with label $\lambda_1$. Call the resulting Euler diagram $d_1$. Trivially, $d_1$ is
    atomic, $H$-extendible and a drawing of $D_1$. Set $i = 1$.
    while $i < n$ do
        Set $\lambda_{i+1}$ to be the label such that $D_i = D_{i+1} - \lambda_{i+1}$.
        if $i == 1$ then
            Add a circle with label $\lambda_{i+1}$ to $d_i$, resulting in $d_{i+1}$.
        else
            Set $ZON = F_i(IN(D_i, D_{i+1}))$, where $IN(D_i, D_{i+1})$ is the set
            $\{az \in Z(D_i) : az \cup \{\lambda_{i+1}\} \in Z(D_{i+1})\}$ and $F_i$ maps abstract zones
            of $D_i$ to zones of $d_i$ in the obvious way.
            Construct $MED(d_i)$.
            if $|ZON| = 2$ and the zones in $ZON$ are adjacent and the edge
            between the zones is in a Hamiltonian cycle in $EGD(d_i)$ or
            $|ZON| = 2$ and we can add two extra zones to $ZON$ to form a valid
            curve-adding cycle or $|ZON| = 3$ and we can add one extra zone to
            $ZON$ to form a valid curve-adding cycle or $|ZON| = 4$ and the zones
            in $ZON$ form a valid curve-adding cycle then
                Add a circle with label $\lambda_{i+1}$ to $d_i$, resulting in $d_{i+1}$.
            else
                Set $s_i$ to be a valid curve-adding cycle for $d_i$ respecting $ZON$,
                selected using the strategy given above.
                Set $d_{i+1} = d_i + (s_i, \lambda_{i+1})$.
            end
        end
        $d_{i+1}$ is atomic and a relaxed drawing of $D_{i+1}$.
        Increment $i$ by 1.
    end
    To finish, set $d = d_n$.
end
```

## 5 Evaluation

We now set out to evaluate the layouts produced by iCurves against competing state-of-the-art techniques. It is not possible to evaluate against all implemented techniques as there are a large number of them. We selected techniques such that (a) the software is freely available, and (b) they could theoretically draw any description, even though the implementation may fail. So, the methods selected for comparison with iCurves are: Bubble Sets [5], GED-L [17], and iCircles [21].

We selected a stratified random sample of 40 data sets from the SNAP (Stanford Network Analysis Project) collection, which are derived from Twitter ego-networks [12]. The 40 data sets had between four and eight sets. The number of zones varied too: for each number of sets, we randomly selected two data sets

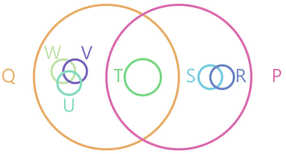


**Fig. 23.** A nested Euler diagram with 8 curves.

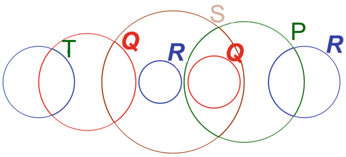


**Fig. 24.** iCircles.

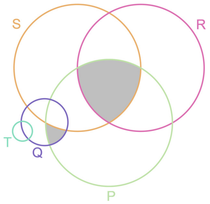


**Fig. 25.** iCurves.

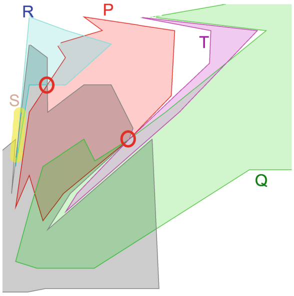


**Fig. 26.** GED-L.

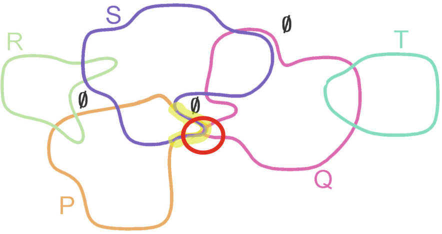


**Fig. 27.** Bubble Sets.

from each quartile based on the number of zones. For example, the eight-set data sets in the SNAP collection had between 8 and 48 zones, with ‘number of zones’ quartiles 8–12, 13–14, 15–18 and 19–48; we randomly chose two data sets from each quartile. This approach to sampling was designed to provide insight into how the effectiveness of the layouts changed as the complexity of the data changed, as measured by the number of sets and the number of zones.

Each selected data set was drawn using each of the four techniques. We counted the number of times each well-formedness property was broken. For each curve label, we counted the number of ‘extra’ times it was used. For each disconnected zone we counted the number of extra minimal regions it comprised (i.e. one fewer than the number of minimal regions of which the zone comprised) on the basis that one minimal region per zone is needed. For concurrency, we counted the number of cases where two or more curve segments run along the same path and we also counted the number of triple points. There were no occurrences of non-simple curves. We also counted the number of non-circular curves. These counts provide a measure of the relative effectiveness of the diagrams. Lastly, we counted the number of extra zones present in the diagrams (i.e. zones whose abstraction is not in the drawn description). Sometimes the techniques failed to draw a given description, so our discussion below accounts for that.

The diagrams generated from the data sets, using all four techniques, are available from https://github.com/AlmasB/d2018. They are marked to show how we counted violations of the well-formedness properties. For example, Figs. 24, 25, 26 and 27 illustrate the same data set drawn using each of the

**Table 1.** Total counts and averages for the diagram features broken by the techniques.

| Techniques | iCircles (40) | | | Bubble Sets (40) | | | GED-L(29) | | | iCurves (38) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $c$ | $a$ | $n$ | $c$ | $a$ | $n$ | $c$ | $a$ | $n$ | $c$ | $a$ | $n$ |
| Duplicated labels | 45 | 1.1 | 17 | | | | | | | | | |
| Disconnected zones | | | | 146 | 3.7 | 21 | 2 | 0.1 | 2 | | | |
| Concurrency | | | | 51 | 1.3 | 12 | 12 | 0.4 | 5 | | | |
| Triple points | | | | 36 | 0.9 | 15 | 20 | 0.7 | 7 | | | |
| Non-circles | | | | 240 | 6.0 | 40 | 46 | 1.6 | 7 | 11 | 0.3 | 8 |
| Extra zones | 5 | 0.1 | 2 | 87 | 2.2 | 20 | | | | 63 | 1.7 | 15 |

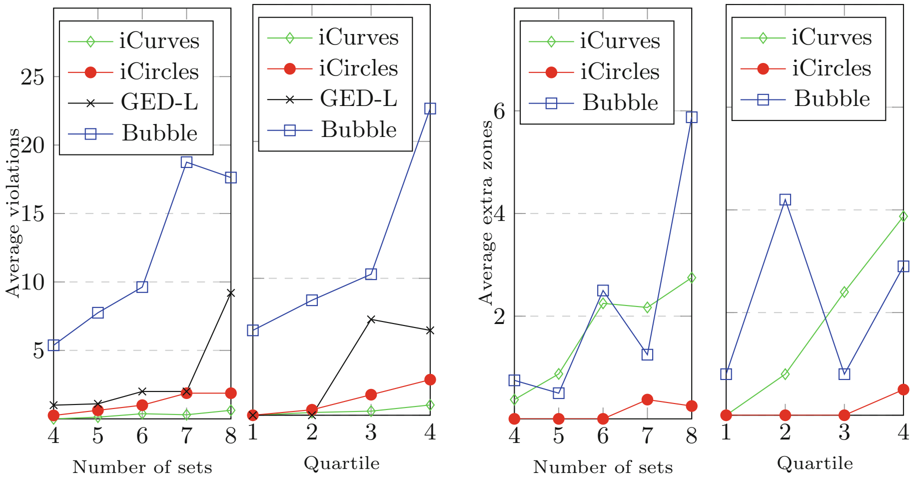


**Fig. 28.** The averages for the diagram features broken by the techniques.

four techniques (the labels are manually added for readability). In Fig. 24 there are two duplicate curve labels. In Fig. 25 there are two extra zones. In Fig. 26, there is one case of concurrency and two triple points. In Fig. 27, there are two cases of disconnected zones, one triple point and one case of curve concurrency.

The data we collected are summarised in Table 1. The column headings show how many of the 40 data sets the technique successfully drew; GED-L failed to generate 11 diagrams out of 40 and iCurves failed to generate 2. In most of the failed drawing attempts the number of zones was high. For each property, the total count, $c$, is given, together with the average number, $a$, of violations for each *drawn* diagram, and the number of diagrams, $n$, in which the violation occurs. When the count is 0, no entry is given to avoid clutter.

The data are also displayed in line charts in Fig. 28. The two charts on the left show the average number of violations of the five well-formedness properties plus the average number of non-circular curves; one graph is by number of sets, the other by quartile. The other two charts show the number of extra zones by number sets and quartile respectively. Bubble Sets often produces diagrams with disconnected zones and extra zones. Furthermore, none of the curves are circles. In iCircles all of the curves are circles, but as the number of zones increases iCircles produces duplicated labels and GED-L produces curve concurrency and triple points. Unsurprisingly, as the diagrams increased in complexity, as indicated by number of sets or zone quartile, the diagrams produced by all four techniques exhibited an increasing number of bad properties.

In summary, we can see that iCircles and Bubble Sets have particularly problematic features, such as duplicated curve labels and, respectively, disconnected zones. These data suggest that GED-L and iCurves strongly outperform iCircles and Bubble Sets based on these two well-formedness properties. GED-L often breaks other well-formedness properties (although not as many times as Bubble Sets) unlike iCurves. Considering effectiveness based just on the well-formedness properties, these data suggest that iCurves should be used rather than GED-L. In addition, iCurves more often used circular curves than GED-L, again pointing favourably towards iCurves. Of course, we must be mindful that the effective features of iCurves come at the expense of extra zones being present.

## 6 Conclusion and Future Work

Our technique, iCurves, generates well-formed diagrams, in contrast with other techniques. The evaluation has shown that iCurves can outperform other techniques with respect to layout features that are detrimental to task performance.

In the future, it would be beneficial to determine the effects of extra zones on diagram comprehension. Understanding the effects would allow us to gain deeper insight into the trade-off between duplicated labels and extra zones, for example in Figs. 24 and 25. Moreover, whilst GED-L did not break any particularly problematic properties, the overall aesthetic qualities of the diagram were questionable. As can be seen from Fig. 26, use of jagged lines makes the diagram convoluted. Currently there are no clear aesthetic criteria we can use to evaluate the diagrams, but properties such as symmetry and curve smoothness can be considered. Finally, we plan to extend iCurves to include graphs, so we can visualize grouped network data (elements joined with lines). Extending the technique in a way that primary spatial rights [5] are not assigned to either the sets or the network represents a considerable challenge.

## References

1. Alper, B., Henry Riche, N., Ramos, G., Czerwinski, M.: Design study of LineSets, a novel set visualization technique. IEEE Trans. Vis. Comput. Graph. **17**(12), 2259–2267 (2011)
2. Alsallakh, B., Micallef, L., Aigner, W., Hauser, H., Miksch, S., Rodgers, P.: Visualizing sets and set-typed data: state-of-the-art and future challenges. In: Eurographics Conference on Visualization STAR, pp. 1–21. Wiley, Hoboken (2014)
3. Blake, A., Stapleton, G., Rodgers, P., Howse, J.: The impact of topological and graphical choices on the perception of Euler diagrams. Inf. Sci. **330**, 455–482 (2016)

4. Chow, S.: Generating and drawing area-proportional euler and venn diagrams. Ph.D. thesis, University of Victoria (2007)
5. Collins, C., Penn, G., Carpendale, M.S.T.: Bubble sets: revealing set relations with isocontours over existing visualizations. IEEE Trans. Vis. Comput. Graph. **15**(6), 1009–1016 (2009)
6. Farrell, G., Sousa, W.: Repeat victimization and hot spots: the overlap and its implication for crime control and problem-oriented policing. Crim. Prev. Stud. **12**, 221–240 (2001)
7. Flower, J., Howse, J.: Generating Euler diagrams. In: Hegarty, M., Meyer, B., Narayanan, N.H. (eds.) Diagrams 2002. LNCS (LNAI), vol. 2317, pp. 61–75. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-46037-3_6
8. Flower, J., Howse, J., Taylor, J.: Nesting in Euler diagrams: syntax, semantics and construction. Softw. Syst. Model. **3**, 55–67 (2004)
9. Gottfried, B.: A comparative study of linear and region based diagrams. J. Spatial Inf. Sci. **2015**(10), 3–20 (2015)
10. Ip, E.: Visualizing multiple regression. J. Stat. Educ. **9**(1) (2001). https://doi.org/10.1080/10691898.2001.11910646
11. Kestler, H., Muller, A., Liu, H., Kane, D., Zeeberg, B., Weinstein, J.: Euler diagrams for visualizing annotated gene expression data. In: Euler Diagrams (2005)
12. Leskovec, J., Krevl, A.: SNAP datasets: Stanford large network dataset collection, June 2014. http://snap.stanford.edu/data. Accessed Sept 2017
13. Pruesse, G., Ruskey, F.: All Simple Venn Diagrams are Hamiltonian. arXiv e-prints, April 2015
14. Riche, N., Dwyer, T.: Untangling Euler diagrams. IEEE Trans. Vis. Comput. Graph. **16**(6), 1090–1099 (2010)
15. Rodgers, P., Zhang, L., Fish, A.: General Euler diagram generation. In: Stapleton, G., Howse, J., Lee, J. (eds.) Diagrams 2008. LNCS (LNAI), vol. 5223, pp. 13–27. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87730-1_6
16. Rodgers, P., Zhang, L., Purchase, H.: Wellformedness properties in Euler diagrams: which should be used? IEEE Trans. Vis. Comput. Graph. **18**(7), 1089–1100 (2012)
17. Rodgers, P.: General embedding method (with diagram library). http://www.eulerdiagrams.com/Library.htm. Accessed Oct 2017
18. Rodgers, P., Stapleton, G., Chapman, P.: Visualizing sets with linear diagrams. ACM Trans. Comput.-Hum. Interact. **22**(6), 27:1–27:39 (2015). http://doi.acm.org/10.1145/2810012
19. Sato, Y., Masuda, S., Someya, Y., Tsujii, T., Watanabe, S.: An fMRI analysis of the efficacy of Euler diagrams in logical reasoning. In: IEEE Symposium on Visual Languages and Human-Centric Computing (2015)
20. Simonetto, P.: Visualisation of overlapping sets and clusters with Euler diagrams. Ph.D. thesis, Université Bordeaux (2012)
21. Stapleton, G., Flower, J., Rodgers, P., Howse, J.: Automatically drawing Euler diagrams with circles. J. Vis. Lang. Comput. **23**(3), 163–193 (2012)
22. Stapleton, G., Rodgers, P., Howse, J.: A general method for drawing area-proportional Euler diagrams. J. Vis. Lang. Comput. **22**(6), 426–442 (2011)
23. Stapleton, G., Rodgers, P., Howse, J., Zhang, L.: Inductively generating Euler diagrams. IEEE Trans. Vis. Comput. Graph. **17**(1), 88–100 (2011)
24. Thièvre, J., Viaud, M., Verroust-Blondet, A.: Using Euler diagrams in traditional library environments. In: Euler Diagrams. ENTCS, vol. 134, pp. 189–202 (2005)
25. Venn, J.: On the diagrammatic and mechanical representation of propositions and reasonings. Philos. Mag. **10**(59), 1–18 (1880)

# Variational Pictures

Martin Erwig and Karl Smeltzer(✉)

Oregon State University, Corvallis, USA
{erwig,smeltzek}@oregonstate.edu

**Abstract.** Diagrams and pictures are a powerful medium to communicate ideas, designs, and art. However, authors of pictures are forced to use rudimentary and ad hoc techniques in managing multiple variants of their creations, such as copying and renaming files or abusing layers in an advanced graphical editing tool. We propose a model of variational pictures as a basis for the design of editors and other tools for managing variation in pictures. This model enjoys a number of theoretical properties that support exploratory graphical design and can help systematize picture creators' workflows.

## 1 Introduction

Visual media such as diagrams and pictures are ubiquitous in the modern world. These take many forms and are employed by a variety of professionals, ranging from architects and industrial designers using CAD tools, graphic designers and photographers using photo editing tools, to scientists and business owners creating charts and technical diagrams to analyze and share data.

While the software tools for these applications are quite sophisticated, they offer little or no support for managing variation in the produced artifacts, forcing users to employ rudimentary techniques to manage multiple versions of their work. For example, a graphic designer who might want to showcase changes to a logo design might be forced to overuse the layer system in their editing tools or simply create multiple copies of the picture file. A data scientist generating a series of similar or related charts and tables might have to manually copy files or images and rename them meaningfully to be able to view and compare them. Architects and engineers frequently make revisions to their drawings and designs and their tools offer minimal, if any support, forcing them to adopt ad hoc approaches.

The need for variation comes in two different forms. First, one might want to create several concurrent variants of an artifact. Second, one might need some kind of version control for pictures. We propose *variational pictures* as an underlying model to support both forms of picture variation. A variational picture is simply a picture (here a grid of pixel values) that offers an explicit way of representing and selecting different variant pictures. For example, a picture

---

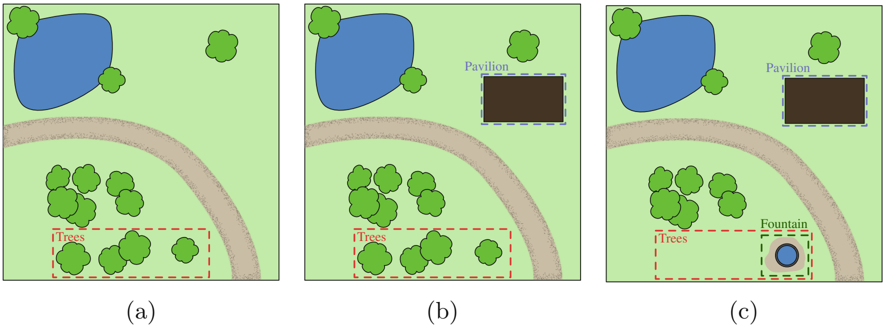
This work is partially supported by the National Science Foundation under the grants IIS-1314384 and CCF-1717300.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 55–70, 2018.
https://doi.org/10.1007/978-3-319-91376-6_9

along with its history (or undo) states can be viewed as a variational picture. Similarly, a collection of related but different designs is also a variational picture. It is even possible to view an animation as a variational picture in which the variants are the frames with a temporal ordering.

Consider a landscape architect working on the plan for a new city park, the features of which are not fully decided yet. For example, an additional wooded area may be cleared to make more space for the park if the budget is determined to allow for it. In another part of the park, a pavilion may be added to provide covered seating. Finally, on the condition that the trees are cleared, a small fountain may be installed where they were. Even with such a small number of undetermined features, this already means there are six possible park layouts. It is easy to imagine this growing out of control rather quickly with more options. By using a model of variational pictures, the architect could produce a single plan with the areas of variability clearly marked that also allows toggling between the options to show them to the final decision makers. We show a sequence of such variational pictures in Fig. 1.

The example does not only show that variational pictures are useful, it also illustrates that managing variability is a nontrivial matter that requires a number of operations for creating, eliminating, and adapting variability. We will return to this example later.



(a) (b) (c)

**Fig. 1.** A sequence of variational pictures showing the design of a park. In (a) we have one area of variability for the potential removal of trees, in (b) we have an independent dimension for a pavilion area, and in (c) we have a nested dimension for a small fountain that can only exist if the trees are removed.

Since variational pictures have a significantly more complicated structure than plain pictures, their precise meaning should be captured in a formal model that can serve as a specification for guiding the implementation of editors and other tools. In particular, a variational picture model should explain the exact behavior of operations for creating, modifying, and navigating variational pictures, as well as for the splitting and merging of variational pictures.

In the following we will present a model for variational pictures and show how it supports editors and other tools for creating and managing variational pictures. Specifically, this work makes the following contributions:

1. A *model of variational pictures* based on the choice calculus [1]. We will present the required background of the choice calculus and how it is used to define a model of variational pictures in Sect. 2.
2. A set of *properties* that describe the generality and usefulness of the presented variational picture model. We will present these properties in Sect. 3.
3. *Requirements* for maintaining variational pictures and a demonstration how the presented model satisfies these requirements. We will discuss the requirements and how they are supported by our variational picture model in Sect. 4.
4. *Variational area trees* to explore and navigate variational pictures diagrammatically, which are presented in Sect. 5.

In Sect. 6 we discuss related work, and we present conclusions and directions for future work in Sect. 7.

## 2 Representing Variational Pictures

The definition of variational pictures is based on a generic model of variational values that is discussed in Sect. 2.1, which we then apply to a model of plain pictures described in Sect. 2.2. The resulting model of variational pictures is then presented in Sect. 2.3. The different degrees of variability in a variational picture give rise to a notion of variation type, introduced in Sect. 2.4, and a corresponding notion of region, which is discussed in Sect. 2.5. Finally, in Sect. 2.6 we introduce an operation that can create variational pictures automatically from a sequence of plain pictures.

### 2.1 A Formal Model of Variation

In order to model and provide structure to variational values we make use of the choice calculus [1], a formal system for managing variation based on the core notion of a choice.

Choices represent sets of *alternatives* associated with names called *dimensions*. For example, we can define a variational integer $A\langle 1, 2\rangle$ as a choice between the two values 1 and 2. In this paper we consider only binary choices, that is, choices of two alternatives. This is not an essential restriction, since choices can be nested to represent a larger number of alternatives. For example, the variational integer $vx = A\langle B\langle 1, 2\rangle, 3\rangle$ contains three total alternatives and has an outer choice in dimension $A$ and a nested choice in dimension $B$. Choice expressions such as this form binary tree structures where $n$ dimension names in the internal nodes lead to $n + 1$ plain values at the leaves.

Each dimension $D$ in a choice expression gives rise to two selectors $D.l$ and $D.r$. Selectors can be used to extract alternatives from choices via a *selection* operation defined as follows (we use $x$ to range over plain, non-variational values, $vl$ and $vr$ to range over (potentially) variational values, and $s$ to range over selectors):

$$\lfloor D\langle vl, vr\rangle\rfloor_s = \begin{cases} \lfloor vl\rfloor_s & \text{if } s = D.l \\ \lfloor vr\rfloor_s & \text{if } s = D.r \\ D\langle\lfloor vl\rfloor_s, \lfloor vr\rfloor_s\rangle & \text{otherwise} \end{cases}$$

$$\lfloor x\rfloor_s = x$$

For example, $\lfloor A\langle 1,2\rangle\rfloor_{A.l} = 1$, $\lfloor vx\rfloor_{A.l} = B\langle 1,2\rangle$ and $\lfloor vx\rfloor_{B.r} = A\langle 2,3\rangle$. Dimensions synchronize choices, which means selecting an alternative in dimension $D$ selects that same alternative in all occurrences of $D$. For example, we have $\lfloor B\langle A\langle 1,2\rangle, A\langle 3,4\rangle\rangle\rfloor_{A.l} = B\langle 1,3\rangle$.

Since one selector can eliminate only choices in one dimension, we generally need a set of selectors, called a *decision*, to extract a plain value from a variational one. This is done by repeated selection with the selectors from the decisions. With $\delta = \{s_1, s_2, \dots, s_n\}$, we have:

$$\lfloor vx\rfloor_\delta = \lfloor\lfloor \dots \lfloor vx\rfloor_{s_1} \dots\rfloor_{s_{n-1}}\rfloor_{s_n}$$

A decision that eliminates all choices from a variational value is said to be *complete*. The order of selection is irrelevant. That is, for any variational value $vx$ and pair of selectors $D.d$ and $D'.d'$ (with $d, d' \in \{l, r\}$):

$$D \neq D' \implies \lfloor\lfloor vx\rfloor_{D'.d'}\rfloor_{D.d} = \lfloor\lfloor vx\rfloor_{D.d}\rfloor_{D'.d'}$$

This will be an important property for variational pictures. As areas of variability will be represented by choices, we can make decisions about optional features in pictures in any order. In the example from the Introduction we could decide, for example, about the fountain before or after we decide about removing the trees.

However, the nesting of the choices *does* matter because it defines dependencies among the decisions to be made. In that same example, having the fountain area nested inside the tree area means that it is possible the outer one will make the inner one irrelevant. This has important consequences for the design of a variational picture editor.

The semantics of a choice calculus expression is given as a mapping from decisions to plain expressions. (This will be made precise in Sect. 2.3.) In general, one variational value can be represented by different choice calculus expressions, which gives rise to a notion of equivalence for expressions that denote the same variational values. For example, choice synchronization is the reason for the choice domination laws that allow the elimination of nested choices in the same dimension:

$$D\langle D\langle vx, vy\rangle, vz\rangle \equiv D\langle vx, vz\rangle \qquad D\langle vx, D\langle vy, vz\rangle\rangle \equiv D\langle vx, vz\rangle$$

Moreover, idempotent choices have no effect on variability:

$$D\langle vx, vx\rangle = vx$$

We also have choice commutation rules, that is, for $D \neq D'$:

$$D'\langle D\langle x, y\rangle, z\rangle \equiv D\langle D'\langle x, z\rangle, D'\langle y, z\rangle\rangle$$
$$D'\langle x, D\langle y, z\rangle\rangle \equiv D\langle D'\langle x, y\rangle, D'\langle x, z\rangle\rangle$$

Although choice commutation preserves the choice calculus semantics, it is still an important consideration for variational pictures, since the nesting of choices relates directly to their structure.

## 2.2 Plain Pictures

We base our model of variational pictures on a model of plain pictures that is basically defined as a set of pixels.

Specifically, given a finite domain of locations *Loc*, a *picture* is a finite mapping from *Loc* to some type $T$. Here $T$ typically is a set of colors, but it can be any type that has an equality predicate defined on it. Moreover, in most cases pictures are given by fixed, rectangular grid of pixels, which means that the type of locations is of the form $Loc^{n,m} = \{1, \ldots, n\} \times \{1, \ldots, m\}$. However, the definitions that follow do not depend on this particular structure, so that we can simply assume a finite set *Loc* of elements on which equality is defined.[1]

Thus the type of $T$ pictures over the domain *Loc* is defined as $Pic_T = Loc \to T$. A picture is an element of that type, and a *pixel* of a picture $p \in Pic_T$ is given by a pair $(l, x) \in p$ with $l \in Loc$ and $x \in T$.

Here is a small example of a $2 \times 2$ picture over a type of symbols $S = \{\circ, \bullet, \star\}$: $p = \begin{smallmatrix}\circ & \bullet \\ \bullet & \circ\end{smallmatrix}$. Since the structure of $T$ doesn't really matter, we will mostly omit it in the following and consider this parameter implicitly fixed, that is, we simply use the type *Pic*.

## 2.3 Adding Choices to Pictures

A *variational picture* of type $T$ is a mapping from locations to variational $T$ values, that is, $VPic_T = Loc \to V(T)$. One could consider a more general definition that also allows variability in the location domain. However, such a type would complicate the following definitions considerably without gaining much. In fact, if the type $T$ contains some unit or null value, one can simulate differently sized pictures using such a designated value. On the other hand, the chosen definition still facilitates the local application of variability, which is an important feature of our model to be discussed later.

Corresponding to plain pictures, a *variational pixel* is an element of a variational picture $vp \in VPic_T$ where $vp = (l, vx)$ with $l \in Loc$ and $vx \in V(T)$. Again, we omit the $T$ parameter from the type in the following.

Here is a variational $2 \times 2$ picture over type $S$ that varies the pixels in $vp$'s second column in dimension $A$: $vp_A = \begin{smallmatrix}\circ & A\langle\bullet,\circ\rangle \\ \bullet & A\langle\circ,\circ\rangle\end{smallmatrix}$. Since $vp_A$ contains only choices

---

[1] This generality follows from the fact that our picture model doesn't require a notion of connectedness.

in one dimension, it encodes two plain variant pictures that can be extracted using selection, that is, $\lfloor vp_A \rfloor_{A.l} = p = \begin{smallmatrix}\circ & \bullet \\ \bullet & \circ\end{smallmatrix}$ and $\lfloor vp_A \rfloor_{A.r} = \begin{smallmatrix}\circ & \circ \\ \bullet & \circ\end{smallmatrix}$.

In general, a variational picture may contain multiple choices in different dimensions. Here is a picture that varies the upper right pixel in $vp_A$ again using a nested $B$ choice in $A$'s left alternative: $vp_{AB} = \begin{smallmatrix}\circ & A\langle B\langle \bullet, \star\rangle, \circ\rangle \\ \bullet & A\langle \circ, \circ\rangle\end{smallmatrix}$. Selecting the left alternative of $A$ in $vp_{AB}$ now does not produce a plain picture, since the $B$ choice has not been eliminated: $\lfloor vp_{AB} \rfloor_{A.l} = \begin{smallmatrix}\circ & B\langle \bullet, \star\rangle \\ \bullet & \circ\end{smallmatrix}$. This means that we need to also perform a selection for $B$. In the example we get $\lfloor vp_{AB} \rfloor_{\{A.l, B.l\}} = p$ and $\lfloor vp_{AB} \rfloor_{\{A.l, B.r\}} = \begin{smallmatrix}\circ & \star \\ \bullet & \circ\end{smallmatrix}$. On the other hand, selecting only the right alternative still produces a plain picture $\lfloor vp_{AB} \rfloor_{A.r} = \begin{smallmatrix}\circ & \circ \\ \bullet & \circ\end{smallmatrix}$.

The semantics of a variational picture is a mapping from decisions to plain pictures. The semantics definition iterates over all variational pixels and extracts and lifts the decisions to the level of pictures, effectively commuting decisions and locations.

$$\begin{aligned} [\![\cdot]\!] &: \mathit{VPic} \to V(\mathit{Pic}) \\ [\![vp]\!] &= \{(\delta, (l, x)) \mid (l, vx) \in vp, (\delta, x) \in vx\} \end{aligned}$$

The type of the semantic function helps explain its definition: since $V(X) = \mathit{Dec} \to X$, $\mathit{Pic} = \mathit{Loc} \to T$, and $\mathit{VPic} = \mathit{Loc} \to V(T)$, the type of the semantic function reads in expanded form as $(\mathit{Loc} \to (\mathit{Dec} \to T)) \to (\mathit{Dec} \to (\mathit{Loc} \to T))$.

As the examples $vp_A$ and $vp_{AB}$ illustrate, the ability to apply choices to individual pixels, makes variability a localized feature. This is an important property, since it allows only those parts to be varied that need it and keeps non-variable picture parts constant across different variant pictures, which supports editing by avoiding update anomalies [2]. For example, if we change the upper left pixel in $vp_{AB}$ from $\circ$ to $\bullet$, this change has to be done only once and will still correctly affect all variants of the picture.

In order to support precise operations to create and modify the variability in a variational picture, we employ the notion of a *view decision*, which is simply a choice calculus decision that specifies the plain picture variant that is currently visible in the editor. View decisions are always complete, that is, they contain exactly one selector for each unique dimension contained anywhere in the variational picture.

### 2.4 Variability Types

The inclusion of choices in pictures suggests a classification of pixels according to their variability and the grouping of pixels with the same variability into regions. To formalize this idea we first define the notion of a *variability type*. The type of a plain, non-variational value is $\diamond$ (called *unit*), and the type of a variational value is given by its choice structure, which is obtained by replacing all plain values in it by $\diamond$. The variability type of a value can be determined by the function $\varphi$, which is defined as follows.

$$\begin{aligned}\varphi &: V(T) \to V(\{\diamond\})\\ \varphi(D\langle vx, vy\rangle) &= D\langle \varphi(vx), \varphi(vy)\rangle\\ \varphi(x) &= \diamond\end{aligned}$$

The type of a pixel is given by the type of its value, that is, $\varphi(l, vx) = (l, \varphi(vx))$.

With this definition, all pixels in $\begin{smallmatrix}\circ & \bullet\\ \bullet & \circ\end{smallmatrix}$ have type $\diamond$. In $vp_{AB} = \begin{smallmatrix}\circ & A\langle B\langle \bullet,\star\rangle,\circ\rangle\\ \bullet & A\langle \circ,\circ\rangle\end{smallmatrix}$, the pixels in the left column have type $\diamond$, the lower right pixel has type $A\langle\diamond,\diamond\rangle$, and the upper right pixel has type $A\langle B\langle\diamond,\diamond\rangle,\diamond\rangle$. For notational convenience we also write more succinctly $A$ for a type $A\langle\diamond,\diamond\rangle$. With this abbreviation we can say that the lower right pixel has type $A$ and the upper right pixel has type $A\langle B,\diamond\rangle$.

A variability type tells us exactly what decisions are needed to extract all plain values from a variational value. For example, the set of plain values contained in $vp_{AB}$ is given by $\{\lfloor vp_{AB}\rfloor_{\{A.l,B.l\}}, \lfloor vp_{AB}\rfloor_{\{A.l,B.r\}}, \lfloor vp_{AB}\rfloor_{A.r}\}$. We can compute the set of decisions required for extracting all plain values from a variational value a particular type with the following function *decs*.

$$\begin{aligned}\mathit{decs} &: V(\{\diamond\}) \to 2^{\mathit{Dec}}\\ \mathit{decs}(D\langle vx, vy\rangle) &= \{\delta \cup \{D.l\} \mid \delta \in \mathit{decs}(vx)\} \cup \{\delta \cup \{D.r\} \mid \delta \in \mathit{decs}(vy)\}\\ \mathit{decs}(x) &= \varnothing\end{aligned}$$

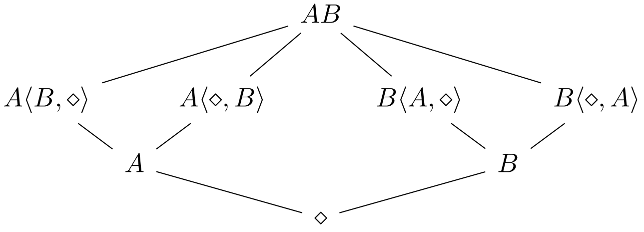
Based on the function *decs* we can define a *variability type equivalence* that holds for types that describe the same variability.

$$\phi \sim \phi' :\Longleftrightarrow \mathit{decs}(\phi) = \mathit{decs}(\phi')$$

Note that $\sim$ is not a simply derived from $\equiv$. For example, whereas $D\langle x, x\rangle \equiv x$ (due to idempotency), $\varphi(D\langle x, x\rangle) = D \not\sim \diamond = \varphi(x)$. We can generalize the notion to type equivalence naturally to a *refinement ordering* on variability types, again based on the decisions that are represented by the variability types.

$$\phi \succsim \phi' :\Longleftrightarrow \forall \delta' \in \mathit{decs}(\phi') : \exists \delta \in \mathit{decs}(\phi) : \delta \supseteq \delta'$$

We have, for example, $A\langle B, B\rangle \succsim A\langle B, \diamond\rangle$ and $A\langle\diamond, B\rangle \succsim A$. The variability refinement is a partial order that give rise to a lattice structure. Here is a small excerpt of this lattice involving the two dimensions $A$ and $B$. Note the two types $A\langle B, B\rangle$ and $B\langle A, A\rangle$ are equivalent. This means that the nesting is not relevant, and we can write the type more accurately as $AB$ to indicate that neither $A$ nor $B$ is in any way privileged over the another.



This diagram indicates that the nesting of dimensions does not matter in fully variationalized values, since the types are equivalent. This is an important property we will come back to in Sect. 3.

We write more shortly $D \in \delta$ whenever $D.l \in \delta$ or $D.r \in \delta$, and say that a dimension $D$ *depends* on dimension $D'$ in a type $\phi$, written as $D' \leftarrow_\phi D$, if $D \in \mathit{decs}(\phi) \implies D' \in \mathit{decs}(\phi)$, that is, in order to make a selection in $D$ one also has to make a selection in $D'$. For example, $B$ depends on $A$ in $A\langle B, \diamond\rangle$. We notice that in the type $AB$ (which is equal to $A\langle B, B\rangle$ and $B\langle A, A\rangle$), $B$ depends on $A$ and $A$ depends on $B$. In such situations, when two dimensions $D$ and $D'$ depend on one another in a type $\phi$, we say that $D$ and $D'$ are *co-dependent* in $\phi$, written as $D' \leftrightarrow_\phi D$.

## 2.5 Variability Regions

Based on variability types we can define a notion of regions that have the same variability. All pixels with the same type have the same variability structure, which means that they are mapped to plain variants by the same set of decisions. This property partitions the set of all pixels (or more precisely, their locations) into a set of disjoint regions. Specifically, for each variation type $\phi \in V(\{\diamond\})$ we define the *region of $\phi$-variability* (or *$\phi$-region* for short) as follows.

$$R_\phi(vp) = \{(l, vx) \in vp \mid \varphi(vx) \sim \phi\}$$

In the park example from the Introduction, the part of the image affiliated with the fountain is given by the region $R_{\mathit{Trees}\langle\diamond,\mathit{Fountain}\rangle}$.

The following lemma is a direct consequence of the definition of $\phi$-regions.

**Lemma 1.** *For every variational picture vp, the set of (non-empty) regions $R_\phi(vp)$ forms a partition of vp.*

Since regions are identified by the common types of their pixels/locations, we can derive a refinement relation for regions based on the type refinement defined earlier. Specifically, $R_{\phi'} \succsim R_\phi$ if and only if $\phi' \succsim \phi$.

As indicated by the example scenario in the Introduction, (variational) pictures are typically the result of a sequence of operations performed in an editor. In particular, variability is introduced into a picture by marking an area and assigning a dimension to it. After that the resulting two alternatives can be edited in different ways and will generally contain different content.

In many cases, however, not every pixel in the marked area will be different in both alternatives. Consider again the picture $vp_A = \begin{smallmatrix} \circ & A\langle\bullet,\circ\rangle \\ \bullet & A\langle\circ,\circ\rangle \end{smallmatrix}$. Both pixels in the right column are variational: they are both of type $A$ and thus belong to the same region. However, only the upper pixel differs in its alternatives. Since both alternatives of the lower pixel are equal, we could apply the idempotency law of the choice calculus and replace $A\langle\circ,\circ\rangle$ by $\circ$ without changing the semantics of the variational picture. Such a change would, however, change the region partition for $vp_A$. By systematically applying transformations for eliminating idempotent choices ($D\langle x, x\rangle \mapsto x$) as well as dominated choices ($D\langle D\langle x, y\rangle, z\rangle \mapsto D\langle x, z\rangle$

and $D\langle x, D\langle y, z\rangle\rangle \mapsto D\langle x, z\rangle)$ to all pixels in a variational picture, we can shrink regions and thus increase the sharing in the picture. We can define a corresponding region shrinking operation as follows. First, we define the operations on variational values. Note that pattern matching on dominated choices is insufficient, since they can occur at arbitrary depths, so we use selection to avoid them instead.

$$\begin{aligned} \dot{\rho}(D\langle vx, vy\rangle) &= \begin{cases} \dot{\rho}(vx) & \text{if } \dot{\rho}(vx) = \dot{\rho}(vy) \\ D\langle \dot{\rho}(\lfloor vx \rfloor_{D.l}), \dot{\rho}(\lfloor vy \rfloor_{D.r})\rangle & \text{otherwise} \end{cases} \\ \dot{\rho}(x) &= x \end{aligned}$$

Region shrinking of variational pictures then simply works by applying the operation $\dot{\rho}$ to all values in all pixels.

$$\begin{gathered} \rho : VPic \to VPic \\ \rho(vp) = \{(l, \dot{\rho}(vx)) \mid (l, vx) \in vp\} \end{gathered}$$

Note that a variational picture obtained by $\rho$ is *not* guaranteed to have maximized the sharing of values and thus is not minimal. Consider, for example, the variational value $vx = A\langle B\langle x, z\rangle, B\langle y, z\rangle\rangle$. The definition of $\dot{\rho}$ cannot extract and share the value $z$, since $\dot{\rho}(vx) = vx$. However, the variational value $vy = B\langle A\langle x, y\rangle, z\rangle$, which is equivalent to $vx$, is smaller than $vx$ and does share $z$.

Note that the redundant value could occur deeply nested in the two alternatives, which means that a simple one-level dimension rotation is, in general, insufficient to expose redundant values. Nevertheless, redundant choices in idempotent or dominated choices will be eliminated by the region shrinking operation. It checks explicitly for identical alternatives and, at every step, uses selection on both branches to ensure no nested choices in matching dimensions.

## 2.6 Distilling Variational Pictures

Given two plain pictures $p$ and $p'$, we can automatically merge them and produce a variational picture that captures the differences between $p$ and $p'$ in choices of some dimension $D$ and keeps all common parts as plain values. First, we define an operation $\dot{\Delta}$ for comparing individual pixel values. If we only needed $\dot{\Delta}$ for generating one variational picture from two plain pictures, its type could be $T \times T \to V(T)$, but since we actually want to apply the merge operation repeatedly to a number of pictures, its type should be $T \times V(T) \to V(T)$, which allows a plain picture to be merged with a variational picture. The operation can, of course, still be used with two plain pictures since a plain picture is just a special case of a variational one.

$$\begin{gathered} \dot{\Delta} : T \times V(T) \to V(T) \\ \dot{\Delta}(x, vx) = \begin{cases} x & \text{if } x = vx \\ D\langle x, vx\rangle & \text{otherwise} \end{cases} \end{gathered}$$

Note that the equality between $x$ and $vx$ can only hold when $vx$ is a plain value. Note also that we have not specified which dimension $D$ is to be used in the definition, since the concrete name does not really matter. We have to postulate however, that $D$ has not been used in any of the $vx$ values that $\dot{\varDelta}$ is applied to. Effectively, we want to use a new dimension for every new picture that is merged into a variational picture.

Using $\dot{\varDelta}$, the operation $\varDelta$ for merging a plain picture into a variational one and capturing their differences in choices can be defined as follows. We assume that both pictures are defined over the same domain of locations. Note that we use the same $l$ in both qualifiers of the set comprehension to express a parallel iteration over all pixels in both pictures.

$$\begin{aligned} \varDelta &: \mathit{Pic} \times \mathit{VPic} \to \mathit{VPic} \\ \varDelta(p, vp) &= \{(l, \dot{\varDelta}(x, vx)) \mid (l, x) \in p, (l, vx) \in vp\} \end{aligned}$$

We can generalize the definition of $\varDelta$ to work on not just two but a whole set of pictures in a straightforward way. The only side condition, which is not formalized here, is that the a fresh dimension is used in each new call to $\dot{\varDelta}$. This could be formalized by threading a set of dimension names through the successive applications of $\varDelta$ and $\dot{\varDelta}$, but this doesn't contribute much to the understanding of the operations, and we therefore omit it here for brevity.

$$\begin{aligned} \varDelta^* &: 2^{\mathit{Pic}} \to \mathit{VPic} \\ \varDelta^*(\{p\}) &= p \\ \varDelta^*(\{p\} \cup P) &= \varDelta(p, \varDelta^*(P)) \end{aligned}$$

We can see this in action using the small plain pictures $p_1 = \begin{smallmatrix}\circ & \bullet \\ \bullet & \circ\end{smallmatrix}$, $p_2 = \begin{smallmatrix}\circ & \circ \\ \bullet & \bullet\end{smallmatrix}$, and $p_3 = \begin{smallmatrix}\circ & \star \\ \bullet & \bullet\end{smallmatrix}$. Now we need to evaluate $\varDelta^*(\{p_1, p_2, p_3\})$, which we can expand to $\varDelta(p_3, \varDelta(p_1, p_2))$. Evaluating $\varDelta(p_1, p_2)$ produces $p_{12} = \begin{smallmatrix}\circ & A\langle\bullet,\circ\rangle \\ \bullet & A\langle\circ,\bullet\rangle\end{smallmatrix}$. Finally, we can evaluate $\varDelta(p_3, p_{12})$ and get the variational picture $\begin{smallmatrix}\circ & B\langle\star, A\langle\bullet,\circ\rangle\rangle \\ \bullet & B\langle\bullet, A\langle\circ,\bullet\rangle\rangle\end{smallmatrix}$.

## 3 Properties of Variational Pictures

In this section we collect a number of general properties of variational pictures that serve as additional characterizations of the concept and also justify the chosen design.

The first observation is that variational picture distillation and semantics are in some sense inverse operations of each other. Distillation is, in fact, what users do when they have a set of pictures in mind that they want to represent in a variational one. Of course, users typically won't encode the differences as efficiently as the operation $\varDelta^*$, which will often result in a variational picture with maximal sharing. Now we can show that the semantics of a variational picture that is generated by $\varDelta^*$ from a set of plain pictures produces exactly the original plain pictures.

**Theorem 1.** $\forall P \in 2^{\mathit{Pic}} : \mathit{range}([\![\varDelta^*(P)]\!]) = P.$

This theorem states that the semantics of variational pictures are correct; it says that distilling a variational picture from a set of plain pictures does not lose any information, because the original set of pictures can be extracted by the semantics.

Since a choice tree with $n$ leaves contains $n-1$ dimensions (as mentioned in Sect. 2.1), that also means that $n$ pictures are distilled into a variational picture with $n-1$ dimensions.

A closely related result is that from the plain pictures encoded in a variational picture we can recover an equivalent variational picture using $\varDelta^*$. Note that the reconstructed variational picture may not be identical to the original one in terms of how the variation is represented, that is, in general we have $\varDelta^*(\mathit{range}(\llbracket vp \rrbracket)) \neq vp$; all we can guarantee is that the reconstructed picture has the same semantics.

**Theorem 2.** $\forall vp \in \mathit{VPic} : \llbracket \varDelta^*(\mathit{range}(\llbracket vp \rrbracket)) \rrbracket = \llbracket vp \rrbracket.$

As the example scenario in the Introduction has shown, areas of variability are often nested, which leads to correspondingly nested choices in the pixel values. This is the case, for example, for the optional fountain. The fountain itself is an area of variability, but it is also nested inside one alternative of another area in which a wooded area is cleared. If we call these dimensions *Trees* and *Fountain*, we would expect to see many pixel values with the type $\mathit{Trees}\langle \diamond, \mathit{Fountain} \rangle$.

One concern is that the initially chosen area and thus choice nesting commits a user to a particular nesting that cannot be changed at a later stage of editing. The next theorem shows that this is actually *not* the case and that variation creation, in particular, the nesting of choices, does not lead to a premature commitment to a particular variation structure.

Consider the situation that an area for choice $B$ has been created inside an area for choice $A$, and let's assume that the $B$ choice lives inside $A$'s left alternative (just as in the example $vp_{AB}$ from Sect. 2.3). The type of the pixels inside the $B$ area is $A\langle B, \diamond \rangle$, and the type of the pixels outside of $B$ but inside of $A$ is simply $A$. We call a pair of regions such as $R_A$ and $R_{A\langle B,\diamond\rangle}$ a *region refinement pair*, since $A\langle B, \diamond \rangle \succsim A$ (and thus $R_{A\langle B,\diamond\rangle} \succsim R_A$), and write such a pair as $R_A[R_{A\langle B,\diamond\rangle}]$ to indicate that it was probably the result of creating a $B$ choice in an $A$ area (or creating an $A$ choice around a $B$ area).[2]

We can observe that for the pixels in $R_{A\langle B,\diamond\rangle}$, $B$ depends on $A$. If we consider the dual case of a region refinement pair $R_B[R_{B\langle A,\diamond\rangle}]$, we can see that the dependency is the other way around, since for the pixels in $R_{B\langle A,\diamond\rangle}$, $A$ depends on $B$. So it seems that the chosen nesting for the areas and dimensions determines two incompatible dependencies among the dimensions. However, there is a straightforward way to reconcile these different refinement pairs by refining the region $R_{A\langle B,\diamond\rangle}$ to $R_{A\langle B,B\rangle} = R_{AB}$ (or dually refining $R_{B\langle A,\diamond\rangle}$ to $R_{B\langle A,A\rangle} = R_{AB}$).

---

[2] Since regions are derived from the pixel types, such region pairs do not necessarily occur in any particular geometric relationship. However, such region pairs typically result from editor actions that create one choice area within another.

A region refinement $R_{A\langle B,\diamond\rangle}$ to $R_{AB}$ can be achieved by simply expanding the right alternative of all $A$ choices in the region's pixels into idempotent $B$ choices, that is, by replacing $A\langle B\langle x,y\rangle,z\rangle$ with $A\langle B\langle x,y\rangle,B\langle z,z\rangle\rangle$.[3]

We can then apply a similar region refinement to $R_A$, turning it as well into a region of type $AB$. This automatically merges the region refinement pair $R_A[R_{A\langle B,\diamond\rangle}]$ into one region $R_{AB}$, and if we assume that the region is not nested within another area, we can consider it together with the surrounding non-variational pixels of type $\diamond$ as a new, bigger region refinement pair $R_\diamond[R_{AB}]$. Next we can refine (part of) the region $R_\diamond$ to $R_B$, leading to $R_B[R_{AB}]$. We can summarize the sequence of transformations as follows.

$$R_A[R_{A\langle B,\diamond\rangle}] \rightsquigarrow R_A[R_{AB}] \rightsquigarrow R_{AB}[R_{AB}] \rightsquigarrow R_{AB} \rightsquigarrow R_\diamond[R_{AB}] \rightsquigarrow R_B[R_{AB}]$$

We can perform a similar transformation for the region pair $R_B[R_{A\langle B,\diamond\rangle}]$:

$$R_B[R_{A\langle B,\diamond\rangle}] \rightsquigarrow R_B[R_{AB}] \rightsquigarrow R_{AB}[R_{AB}] \rightsquigarrow R_{AB} \rightsquigarrow R_\diamond[R_{AB}] \rightsquigarrow R_A[R_{AB}]$$

This shows that while we can't turn $R_A[R_{A\langle B,\diamond\rangle}]$ into $R_B[R_{A\langle B,\diamond\rangle}]$ (or vice versa) without removing information, we can invert the nesting of choices if we refine the innermost region sufficiently.

**Theorem 3.** *$R_A[R_{A\langle B,\diamond\rangle}] \rightsquigarrow^* R_B[R_{AB}]$ and $R_B[R_{A\langle B,\diamond\rangle}] \rightsquigarrow^* R_A[R_{AB}]$*

We consider an application of this theorem in the next section.

## 4 Maintenance of Variational Pictures

It is unrealistic to expect variational picture authors to know the precise and final locations of variability they will need from the outset. This means that our model of pictures should support changing areas of variability that have already been defined.

Consider the park example from the Introduction. Suppose that, after creating the design shown, the architect is told that the *Fountain* area needs to include pipes connecting from the water main. This means that the associated area needs to not only grow, but grow such that the nesting order with the *Trees* area is reversed. The final result can be seen in Fig. 2.

Fortunately, we have already seen in Sect. 3 that region refinement allows us to change the nesting order without issue. We just need to transform the appropriate pixels by expanding them all with an outer *Fountain* choice. There are three transformations to make, namely for those pixels originally non-variational which are now contained in *Fountain*, those which were originally in *Trees* but outside of *Fountain*, and those inside of *Fountain*. They are transformed as follows.

---

[3] There are situations in which other transformations, such as $A\langle B\langle x,y\rangle,B\langle z,y\rangle\rangle$ may be more appropriate, but the point is that a region refinement is easy to achieve.

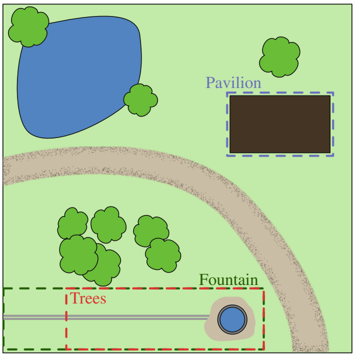


**Fig. 2.** The variational picture after resizing the previously nested *Fountain* area to contain the *Trees* area, in order to depict the connecting water pipes.

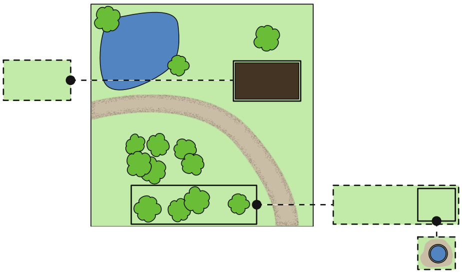


**Fig. 3.** A Variational Area Tree (VAT) that showing an entire variational picture at once. Here the VAT for the view decision $\{\mathit{Trees}.l, \mathit{Fountain}.l, \mathit{Pavilion}.r\}$ is shown.

$$\begin{aligned} x &\mapsto \mathit{Fountain}\langle x, x\rangle \\ \mathit{Trees}\langle x, y\rangle &\mapsto \mathit{Fountain}\langle \mathit{Trees}\langle x, y\rangle, \mathit{Trees}\langle x, y\rangle\rangle \\ \mathit{Trees}\langle x, \mathit{Fountain}\langle y, z\rangle\rangle &\mapsto \mathit{Fountain}\langle \mathit{Trees}\langle x, y\rangle, \mathit{Trees}\langle x, z\rangle\rangle \end{aligned}$$

Although we do not depict it here, we could similarly envision scenarios where we want to shrink regions. The swapping works in exactly the opposite way to expanding. The only difference is we must remove part of the variability by performing a selection of the dimension that we are shrinking. Since we need to choose one alternative or the other to select, we defer to the value of the view decision. This gives the user control over the different possibilities.

## 5 Variational Area Trees

Although the described model of variational pictures is sufficiently flexible to build variational pictures, so far we have limited ourselves to viewing a single variant at a time. While this simplifies editing operations, there is still a need to produce overviews to better understand and navigate the pictures, which calls for a visual language with a *graphical* domain [3]. To this end, we present *variational area trees* (VATs), a diagrammatic approach to viewing an entire variational pictures simultaneously.

VATs show the currently selected variant in its entirety as a root node, and then also all of the other possible selections as branches. Left and right alternatives are always connected via a dotted line and shown either to the left and right of one another or above and below, in order to use space more efficiently (we assume a reasonable layout algorithm). Dotted outlines show unselected variants and solid outlines indicate selected ones. A variational area tree for the view decision $\{Trees.l, Fountain.l, Pavilion.r\}$ the park example are shown in Fig. 3.

VATs have a number of use cases. Obviously, they provide a concise overview of an entire variational pictures including all of its variants. In addition, the design of VATs has a number of useful properties.

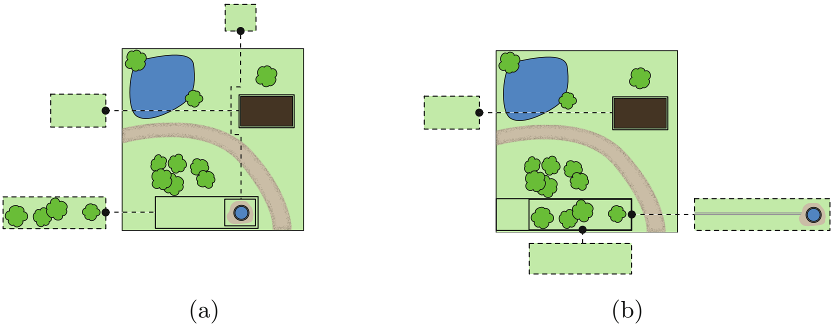

(a) (b)

**Fig. 4.** Additional configurations of the park picture VAT. In (a) we see the view decision $\{Trees.r, Fountain.r, Pavilion.r\}$, and (b) shows the case after the commuting of the *Trees* and *Fountain* regions for the view decision $\{Trees.l, Fountain.l, Pavilion.r\}$.

First, the total number of regions shown gives the total number of different drawing areas and provides a clue to the variability of the picture. Second, since unselected areas are never nested and thus all unselected areas are always placed on the top level, counting all top-level unselected areas provides a concrete measure of what is hidden in the current view. Third, the number of boundaries that are crossed by a (horizontal or vertical) connector line indicates the nesting level of that choice/variational area, and the types of the crossed boundaries (unselected vs. selected) tell immediately what decisions have to be flipped (at minimum) to make the area visible. Fourth, VATs illustrate nicely where in the choice tree the current variant is located. For example, the VAT in Fig. 4(a) is located rightmost/bottommost, which means that the right alternative must be selected in all dimensions.

VATs can also serve as quick navigation tools. It is easy to conceive of an interface in which a user can zoom out to a VAT and then change the selection of arbitrary dimensions quickly. Finally, VATs could also support compound operations or queries. For example, a picture creator may want to view the parts of the picture that are not variational or all the areas affected by a particular dimension. These kinds of operations could be supported by a simple filtering operation on VATs. Figure 4 demonstrates some additional examples for different selections and after expanding the *Fountain* region as done in Sect. 4.

## 6 Related Work

Although there are a large number of potential applications for variational pictures, the research in this area is limited in scope. There are many tools design to offer digital image version control and asset management, although most are proprietary and do not describe their specific model. Examples of these include Adobe Drive and AutoDesk Vault. Some more general version control tools offer support for image diff operations including Perforce and Git via services such as Github.

This topic also emerges in the field of information visualization, Heer et al. [4] performed a large survey of graphical history tools from the context of information visualization and exploratory analysis which covers this topic well beyond the scope of our work. Chen et al. [5] proposed a graph-based revision control system for images. Being based on graphs, it focuses on paths of editing operations rather than pixels or image objects, and challenges such as diff and merge are solved with graph operations. They also offer support for selective undo and "nonlinear exploration", in which the user can adjust parameters to operations that have already been performed. Gleicher et al. [6] demonstrated comparative visualization as a fundamental idea, which can be viewed as an application of variability to pictures within the scope of information visualization. To our knowledge, none of this work describes a general variability-aware model either.

Another related body of work is on model difference techniques. Kolovos et al. [7] includes an overview of the topic. Specific examples include Ohst et al. [8] who described an approach to model difference in UML and Cicchetti et al. [9] who showed a technique for representing differences between models independent of the metamodel. As models are generally expressed using graphs, none of this work describes a pixel-based approach.

Terry and Mynatt proposed Side Views [10] for previewing commands such as image rotation and coordinate transformations, which makes use of variational pictures but does not model them explicitly. Terry et al. [11] expanded on this by managing variations more explicitly. Their tool, Parallel Paths, allows operations to be applied to individual picture variants or groups of them and also tracks history to navigate throughout their notion of a variational picture. That work is primarily focused on a specific user interface, however, and not on a formal model of variational pictures. Hartmann et al. [12] described an approach to creating variational interactions and user interfaces based on editing linked source code alternatives and parallel execution. Finally, Foo et al. [13] summarize existing work and propose new ideas in the challenge of identifying similar (in the sense of different resolutions, compression techniques, fragments, etc.) images. This work focuses on identifying similar images rather than explicitly managing the variability.

## 7 Conclusions and Future Work

In this work we have introduced the notion of a variational pictures and introduced a formal model that captures their functionality and behavior based on the choice calculus. We have established several basic properties of the model that support tools for editing and transforming variational pictures. The generality of the presented model (as reflected, for example, by the general types *Loc* and $T$) allows more specific models of variational pictures to be targeted at specific application domains, which provides a rich area for future work.

## References

1. Erwig, M., Walkingshaw, E.: The choice calculus: a representation for software variation. ACM Trans. Soft. Eng. Methodol. **21**(1), 6:1–6:27 (2011)
2. Date, C.J.: Database in Depth: Relational Theory for Practitioners. O'Reilly Media Inc., Sebastopol (2005)
3. Erwig, M., Smeltzer, K., Wang, X.: What is a visual language? J. Vis. Lang. Comput. **38**(C), 9–17 (2017)
4. Heer, J., Mackinlay, J., Stole, C., Agrawala, M.: Graphical histories for visualization: supporting analysis, communication, and evaluation. IEEE Trans. Vis. Comput. Graph. **14**, 1189–1196 (2008)
5. Chen, H.T., Wei, L.Y., Chang, C.F.: Nonlinear revision control for images. In: ACM SIGGRAPH, pp. 105:1–105:10 (2011)
6. Gleicher, M., Albers, D., Walker, R., Ilir, J., Hansen, C.D., Robers, J.C.: Visual comparison for information visualization. Inf. Vis. **10**, 289–309 (2011)
7. Kolovos, D.S., Ruscio, D.D., Pierantonio, A., Paige, R.F.: Different models for model matching: an analysis of approaches to support model differencing. In: ICSE Workshop on Comparison and Versioning of Software Models, pp. 1–6 (2009)
8. Ohst, D., Welle, M., Kelter, U.: Differences between versions of UML diagrams. In: European Software Engineering Conference Held Jointly with ACM SIGSOFT International Symposium on Foundations of Software Engineering, pp. 227–236 (2003)
9. Cicchetti, A., Di Ruscio, D., Pierantonio, A.: A metamodel independent approach to difference representation. J. Object Technol. **6**(9), 165–185 (2007)
10. Terry, M., Mynatt, E.D.: Side views: persistent, on-demand previews for open-ended tasks. In: ACM Symposium on User Interface Software and Technology, pp. 71–80 (2002)
11. Terry, M., Mynatt, E.D., Nakakoji, K., Yamamoto, Y.: Variation in element and action: supporting simultaneous development of alternative solutions. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 711–718 (2004)
12. Hartmann, B., Yu, L., Allison, A., Yang, Y., Klemmer, S.R.: Design as exploration: creating interface alternatives through parallel authoring and runtime tuning. In: ACM Symposium on User Interface Software and Technology, pp. 91–100 (2008)
13. Foo, J.J., Sinha, R., Zobel, J.: Discovery of image versions in large collections. In: Cham, T.-J., Cai, J., Dorai, C., Rajan, D., Chua, T.-S., Chia, L.-T. (eds.) MMM 2007. LNCS, vol. 4352, pp. 433–442. Springer, Heidelberg (2006). https://doi.org/10.1007/978-3-540-69429-8_44

# Edge Label Placement in Layered Graph Drawing

Christoph Daniel Schulze(✉), Nis Wechselberg, and Reinhard von Hanxleden

Department of Computer Science, Kiel University, Kiel, Germany
`{cds,nbw,rvh}@informatik.uni-kiel.de`

**Abstract.** Many visual languages based on node-link diagrams use edge labels. We describe different strategies of placing edge labels in the context of the layered approach to graph drawing and investigate ways of encoding edge direction in labels. We also report on the results of experiments conducted to investigate the effectiveness of the strategies.

## 1 Introduction

Visual programming languages based on *node-link diagrams*, such as Sequentially Constructive Charts (SCCharts) [13] (a synchronous state charts dialect, see Fig. 1), have become mainstream in several industries. Many share a number of similarities: first, being based on a notion of either data flow (data is produced, processed, and consumed by *nodes* and transmitted between them through *links* or *edges*) or control flow (nodes represent *states* that can be active or not, with *transitions* transferring control between them); second, deriving some of their semantics through textual labels; and third, requiring users to spend a considerable amount of time on laying out their diagrams [7] for them to properly readable [9], giving rise to automatic layout algorithms [12].

A popular layout approach for flow-based diagrams is the layered approach introduced by Sugiyama et al. [11], which tends to emphasize data or control flow by making the majority of edges point in the same direction. The original description of the layered approach did not mention edge labels. Not taking them into account, however, will lead to layouts with too little space available for their placement, resulting in overlaps with other diagram elements—something that may well cause users to refrain from using automatic layout in the first place. This paper is about making labels first-class citizens during automatic layout.

Due to its prevalence in flow-based diagrams, we will limit our discussions to horizontal layout directions. While the question of how well-suited the methods are to vertical layouts is interesting, it is outside the scope of this paper due to space constraints.

**Contributions.** We show different ways of placing labels within the layered approach, including the selection of layers to place labels in and the side of their edge to place them on. We also investigate ways of encoding an edge's direction

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 71–78, 2018.
https://doi.org/10.1007/978-3-319-91376-6_10

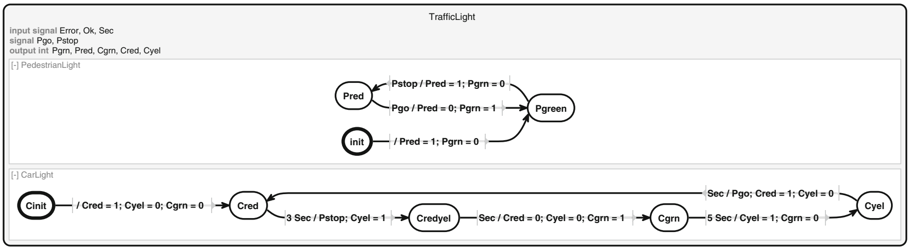


**Fig. 1.** An SCChart laid out with the methods we propose in this paper. This drawing uses a horizontal layout with edges routed as splines.

through label placement or additional decorations, intended to be of particular help in use cases where only parts of a diagram can be displayed on screen. We summarize results of a controlled experiment as to the effectiveness of the proposed strategies.

**Related Work.** Label placement in general has a long history in cartography. In a classic paper [4], Imhof lays down six principles for good map labeling, which Kakoulis and Tollis [5] apply to edge labeling as the following three rules:

1. No overlaps between labels and other diagram elements.
2. It should be clear which diagram element a label belongs to. Imhof calls this "clear graphic association".
3. Among all acceptable positions, a label should be placed in the best possible.

Kakoulis and Tollis also provide a definition of the *edge label placement problem*, which is about placing edge labels in diagrams whose elements have already been placed. Existing algorithms, of which Kakoulis and Tollis provide an overview [12, Chap. 15], usually either run the risk of violating rules 1 or 2 or may resort to hiding or at least scaling down labels to avoid violations—both undesirable for visual programming languages.

In this paper we consider label placement a part of automatic layout, thereby ensuring that there will always be enough space available to satisfy rules 1 and 2. Klau and Mutzel [6] do this for the topology-shape-metrics approach to graph drawing, although their results do not always seem to satisfy rule 2. The *Graphviz dot*[1] algorithm, an implementation of the layered approach, handles edge labels by introducing dummy nodes [2], an approach we follow as well. However, they do not describe any strategies regarding where edge labels end up with regard to their edge. Castelló et al. [1] place labels on edges, which is also one of our label placement strategies. However, they do not discuss graphical design considerations and do not evaluate whether doing so may have a negative impact on the ability of users to read the resulting drawing.

---

[1] http://www.graphviz.org/.

There have been more radical proposals, most notably by Wong et al. [15] who replace an edge by its label. That approach would not work with long edges or orthogonal edge routing, but our on-edge label placement strategy to be introduced in Sect. 3 can be seen as a less extreme version of this technique.

Different methods have been proposed to indicate edge direction, such as using curvature, or color and thickness gradients from an edge's tail to its head [3,16]. These will cease to work in use cases where users only see a small part of a larger diagram and changes in such features are subtle. Animating edges or rendering them as sequences of arrows [3], may work well, but increase visual clutter and require the rendering of edges to be changed, which may be impossible if it carries semantical meaning (as, for example, it does in *LabVIEW* by *National Instruments*). Our methods do not require such design changes.

**Outline.** We describe label placement techniques in Sects. 2 and 3, respectively, before introducing directional decorators in Sect. 4. We evaluate the techniques in Sect. 5 and conclude in Sect. 6.

An extended version of this paper that includes detailed descriptions of the experiments we report on is available as a technical report [10].

## 2 Layer Selection

The layered approach is split into five phases. *Cycle breaking* reverses edges in the input graph to make the graph acyclic, to be restored again once the algorithm has finished. *Layer assignment* partitions the set of nodes into a sequence of *layers* such that edges only point to layers further down the sequence. Edges that span multiple layers are broken by introducing *dummy nodes* such that edges always connect nodes in adjacent layers. *Crossing reduction* orders the nodes in each layer to reduce the number of edge crossings. *Node placement* computes vertical node coordinates and thereby determines the height of the diagram. *Edge routing* finally computes bend points for the edges according to the preferred edge routing style. For flow-based diagrams, this will usually be orthogonal edge routing or spline edge routing. The methods to be described in this paper work with both.

The aim of integrating edge label placement into the layout algorithm is to reserve enough space for the edge labels to be placed without overlaps and with clear graphic association. Similar to Graphviz dot, we break each edge that has labels by introducing a *label dummy node* to represent them. We compute the size of the dummy node such that all edge labels fit into it, stacked upon each other with a configurable amount of space between them, plus spacing to be left between the labels and their edge. Once edge routing has finished, label dummies are replaced by the labels they represent.

Label dummies need to be inserted before the layer assignment step to ensure that each dummy is assigned to a layer (which might end up existing only because of the label dummy). After layer assignment, we can move each label dummy to a layer of our choice, if necessary. That choice is obvious if the edge is so short that there is only one layer to choose from. If the edge is longer, however—such as the edge from `cyel` to `cred` in Fig. 1—we need a strategy that defines what constitutes the best choice.

If the edge spans layers $L_1$ through $L_n$, two simple strategies are obvious. The *median strategy* places the label dummy in layer $L_{\lfloor n/2 \rfloor}$, while the *end layer strategy* places it either in layer $L_1$ (*source layer strategy*) or $L_n$ (*target layer strategy*).

The most appropriate strategy depends on the visual language. In SCCharts, for example, edges represent transitions from a source to a target state that are eligible to be taken based on some condition, which is part of each transition's label. If edge labels are placed using the median strategy, a user might have to search a large area of an SCChart in order to understand a single transition. In this case, the source layer strategy may be more helpful.

An optimization goal might be to assign labels to layers such that the drawing's width is minimized. While taking layer widths into account seems easy enough, it is complicated by the fact that the widths may be changed by the assignment itself. Finding good algorithms to solve this problem is the subject of ongoing research and transcends the scope of this paper.

## 3 Label Side Selection

An edge label can be placed above, below, or even on the edge it belongs to, and we can think of different strategies to make a decision.

### 3.1 Same-Side Strategy

The *same-side strategy* places all labels either above or below their edge, as shown in Fig. 2a. The simplest strategy to implement, it may also be the easiest for users to understand due to its consistency.

For clear graphic association to be achieved, it does require the spacing between a label and its edge to be noticeably smaller than the spacing between the label and other edges, in accordance with the Gestalt principle of *perceptual grouping* [14]. Donald Norman would call this "knowledge in the world" [8] in that users do not require further information to decipher the diagram.

If spacings are chosen badly, the same-side strategy can still work if the user is aware of it. Norman, however, claims that such additional information required to understand the world—what he calls "knowledge in the head"—should be avoided when possible, making properly chosen spacings the preferred method.

### 3.2 Directional Strategy

The same-side strategy works well in terms of clear graphic association, but does not encode the direction an edge is heading towards. Since a label may be far removed from the end points of its edge (which, depending on the graph's size and the way it is displayed, may not even be on screen), any clue as to the edge

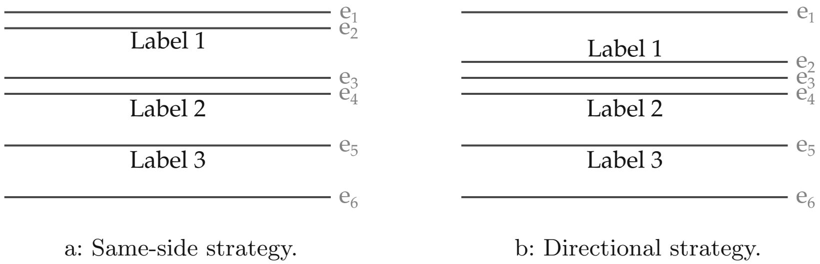


**Fig. 2.** Two different label side selection strategies. Both place labels closer to their edge than to other edges. Note that this is only an excerpt of a graph, which explains the absence of edge arrows or nodes.

direction may help a user navigate the diagram. The *directional strategy* aims to do just that by placing labels above edges running rightwards and below edges running leftwards.

Figure 2b shows an example of this strategy in action. Knowing about this placement method lets us deduce that $e_2$ is headed rightwards while $e_4$ and $e_5$ are going off to the left. If spacings are chosen well, this additional piece of knowledge is not required for clear graphic association, but offers clues to advanced users of a visual language who know about the convention. If spacings are chosen badly, however, the directional strategy ceases to work due to the ambiguity it would produce.

Of course, this strategy requires knowledge in the head, which we will improve upon in Sect. 4.

### 3.3 On-Edge Strategy

We have thus far focused on placing labels next to their edge, which is the standard edge labeling strategy in the vast majority of graphical modeling tools, such as *Papyrus* (Eclipse Foundation) or *Simulink* (MathWorks). There is a case to be made, however, for placing them *on* their edge.

When placing labels next to their edge, one of our main concerns has to be clear graphic association. Wong et al. [15] respond to that challenge by replacing the edge with its label. We will not follow their proposal, for several reasons. First, for the approach to work without introducing distortion or very different font sizes, the length of an edge would have to be a function of the text it is labeled with—a prerequisite not compatible with the layered approach. And second, the orthogonal edge routing style (or any routing style that employs bend points, for that matter) would degrade label legibility even further.

On-edge label placement achieves optimal graphic association without completely replacing edges by their label. If the layout direction is horizontal, we may also reduce the diagram's height slightly because there is no edge-label spacing anymore, and since each label sits on its edge the space between the

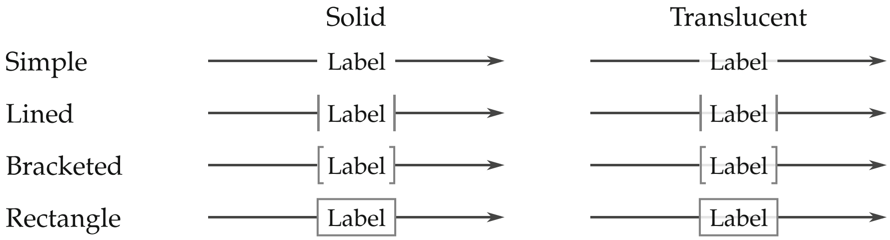


**Fig. 3.** Four examples on-edge label designs.

label and unrelated edges can be much smaller than it could otherwise. For on-edge label placement to work, the graphical representation of edge labels has to be designed accordingly. Labels must have either a solid background or at least cause the background to be sufficiently faded for the edge not to interfere with the text's legibility. This requirement is easy to meet, and many designs for on-edge labels are possible, which may even reflect different edge semantics. Figure 3 shows four simple examples of on-edge label representations. Castelló et al. [1] use a simple solid design when drawing statecharts, but do not discuss their motivation for doing so. Interrupting the edge, however, may cause users to have a harder time following it through the diagram.

## 4 Directional Decorators

The directional label side selection strategy had the advantage of encoding information about the direction of an edge, but suffered from both potential graphic association problems as well as knowledge in the head for its proper interpretation. An alternative is to communicate through the label's design instead of its placement. Figure 4 shows examples of on-edge labels decorated with an additional arrow which points towards the edge's head. Such decorations work with any label side selection strategy, thus allowing the same-side strategy to communicate the same amount of information as the directional strategy while being slightly clearer in terms of graphic association.

**Fig. 4.** Labels can be decorated with arrows to point at where the edge is heading. While this example only shows on-edge labels, such decorations can of course also be added to labels placed next to their edge.

An interesting problem concerns the implementation of directional decorators. Whether the arrow should point leftwards or rightwards is subject to the diagram's layout, which implies that the viewing framework needs to support changes to the visualization after automatic layout has run. How this can be done depends on the viewing framework and is outside the scope of this paper.

## 5 Evaluation

We conducted a controlled within-participants experiment with 48 participants in order to answer two research questions (a detailed account of the experiment is available in the expanded technical report [10]).

First, are users better at inferring edge directions with directional label placement or with on-edge label placement with directional decorators? We showed users random images with lines labeled using one of the two strategies, intended to simulate seeing excerpts of larger diagrams. We found that they had a significantly faster response time and significantly lower error rate with on-edge labels.

Second, does on-edge label placement have a negative impact on the ability of users to follow edges through a diagram? We showed users graphs that had a start node highlighted and asked how many nodes were reachable from that node in two steps. Among three conditions (same-side, directional, and on-edge) we could not find significant differences in response time or error rate.

In a concluding interview we asked participants to rank four label placement strategies (same-side, directional, on-edge without and with directional decorators). The latter was ranked significantly higher than the other three strategies, among which we did not find significant differences. The directional strategy was often described as being confusing. Some participants mentioned that the value of on-edge label placement with directional decorators increases with a diagram's size, noting that the additional arrows can add visual clutter to small diagrams.

## 6 Conclusion

We presented different placement strategies for placing edge labels in flow-based diagrams. On-edge label placement yields clearest graphical association, and usually slightly smaller diagrams. With directional decorators added it was largely preferred by users. Some did complain about the fact that on-edge labels interrupt their edges, but we did not find significant performance differences in a task that required participants to follow edges through a diagram.

**Future Work.** Some visual languages tend to produce rather long edge labels that make the assignment of labels to layers a crucial influence on the width of drawings. Finding a heuristic that yields assignments that produce smaller drawings seems necessary.

Clear graphic association of on-edge labels may be impaired if labels span multiple lines of text. This issue should be investigated further.

**Acknowledgements.** The authors would like to thank Nadine Yarar, Andreas Mühling, and Emmanuel Manalo for helpful suggestions while designing the experiment, and Helen Purchase for her excellent book on the subject.

This work has been supported in part by the German Science Foundation, as part of the Compact Graph Drawing with Port Constraints (ComDraPor, DFG HA 4407/8-1) project.

## References

1. Castelló, R., Mili, R., Tollis, I.G.: An algorithmic framework for visualizing statecharts. In: Marks, J. (ed.) GD 2000. LNCS, vol. 1984, pp. 139–149. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44541-2_13
2. Gansner, E.R., Koutsofios, E., North, S.C., Vo, K.-P.: A technique for drawing directed graphs. Softw. Eng. **19**(3), 214–230 (1993)
3. Holten, D., Isenberg, P., van Wijk, J.J., Fekete, J.-D.: An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs. In: 2011 IEEE Pacific Visualization Symposium, pp. 195–202, March 2011
4. Imhof, E.: Positioning names on maps. Am. Cartogr. **2**(2), 128–144 (1975)
5. Kakoulis, K.G., Tollis, I.G.: An algorithm for labeling edges of hierarchical drawings. In: Di Battista, G. (ed.) GD 1997. LNCS, vol. 1353, pp. 169–180. Springer, Heidelberg (1997). https://doi.org/10.1007/3-540-63938-1_60
6. Klau, G.W., Mutzel, P.: Combining graph labeling and compaction. In: Kratochvíyl, J. (ed.) GD 1999. LNCS, vol. 1731, pp. 27–37. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-46648-7_3
7. Klauske, L.K., Dziobek, C.: Improving modeling usability: automated layout generation for Simulink. In: Proceedings of the MathWorks Automotive Conference (MAC 2010) (2010)
8. Norman, D.A.: The Design of Everyday Things. Basic Books, New York (1988)
9. Petre, M.: Why looking isn't always seeing: readership skills and graphical programming. Commun. ACM **38**(6), 33–44 (1995)
10. Schulze, C.D., Wechselberg, N., von Hanxleden, R.: Edge label placement in layered graph drawing. Technical report 1802, Department of Computer Science, Christian-Albrechts-Universität zu Kiel, February 2018. ISSN 2192-6247
11. Sugiyama, K., Tagawa, S., Toda, M.: Methods for visual understanding of hierarchical system structures. IEEE Trans. Syst. Man Cybern. **11**(2), 109–125 (1981)
12. Tamassia, R. (ed.): Handbook of Graph Drawing and Visualization. CRC Press, Boca Raton (2013)
13. von Hanxleden, R., Mendler, M., Aguado, J., Duderstadt, B., Fuhrmann, I., Motika, C., Mercer, S., O'Brien, O., Roop, P.: Sequentially constructive concurrency—a conservative extension of the synchronous model of computation. ACM Trans. Embed. Comput. Syst. Spec. Issue Appl. Concurr. Syst. Des. **13**(4s), 144:1–144:26 (2014)
14. Wagemans, J., Elder, J.H., Kubovy, M., Palmer, S.E., Peterson, M.A., Singh, M., von der Heydt, R.: A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. Psychol. Bull. **138**(6), 1172–1217 (2012)
15. Wong, P.C., Mackey, P., Perrine, K., Eagan, J., Foote, H., Thomas, J.: Dynamic visualization of graphs with extended labels. In: Proceedings of the 2005 IEEE Symposium on Information Visualization, INFOVIS 2005, p. 10. IEEE Computer Society, Washington, D.C. (2005)
16. Xu, K., Rooney, C., Passmore, P., Ham, D.-H., Nguyen, P.H.: A user study on curved edges in graph visualization. IEEE Trans. Vis. Comput. Graph. **18**(12), 2449–2456 (2012)

# Generation of Kolam-Designs Based on Contextual Array P Systems

Ibrahim Venkat[1], Thamburaj Robinson[2], K. G. Subramanian[2](✉), and Philippe de Wilde[3]

[1] School of Computer Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia
`ibra@usm.my`
[2] Department of Mathematics, Madras Christian College,
Tambaram, Chennai 600059, India
`robinson@mcc.edu.in`, `kgsmani1948@gmail.com`
[3] University of Kent, Canterbury, UK
`p.dewilde@kent.ac.uk`

**Abstract.** Kolam-designs are diagrams used to decorate the floor, especially in front of a house in South India. Methods of generation of the kolam diagrams were developed based on two-dimensional picture generating models, broadly known as array grammars, introduced for the description and analysis of picture patterns. Rewriting array P system, a membrane computing model based on array rewriting has been developed to evolve picture arrays, based on context-free array rewriting rules. In contrast to this array P system, another P system model called contextual array P system (CAP) using contextual array rules for the evolution or generation of picture arrays has been proposed and its power in generating picture arrays investigated. Here we develop an application of CAP for the generation of the kolam diagrams. The advantage of using CAP is that kolam diagrams that cannot be handled by array grammars can be generated by the CAP model.

**Keywords:** Floor-designs · Kolam patterns · Array grammars
Array P system · Contextual array rules

## 1 Introduction

Kolam-designs are visually pleasing geometric diagrams that are drawn in the ground mostly at the entrance of a home [1,5], with this age-old tradition being more prevalent in South India. A kolam is generally drawn starting with a certain number of points arranged in a particular pattern, with curly lines being drawn around these points, resulting in the intended "kolam" drawing. In fact there are very intricate and complicated kolam patterns used by the kolam practitioners. Motivated by these "kolam" floor-designs Siromoney et al. [8], in the years 1972–74, proposed picture generating models, called array grammars, and described a technique [8], referred to as Narasimhan's method, for the generation

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 79–86, 2018.
https://doi.org/10.1007/978-3-319-91376-6_11

of kolam patterns using these array grammars. Subsequently, several researchers have developed a variety of picture-array models. These grammars make use of two-dimensional extension of Chomsky type of string grammars or contextual type of string grammars [7] in the area of formal languages. On the other hand a bio-inspired computing model with a generic name of P system was introduced by Paun [6] which has turned out to be a versatile model with applications in different areas, with the area of picture array generation being one among these application areas. Several P system models have been proposed for picture array generation [2,9]. Based on contextual type array generating rules [4], an array P system, called contextual array P system ($CAP$) has been introduced in [3]. Here we construct $CAP$ for generation of kolam patterns that cannot be handled by contextual array grammars [4] utilizing the technique of Narasimhan's method of kolam generation.

## 2 Preliminaries

We recall needed notions and related results [3,4]. A point $p = (m, n)$ in the two-dimensional digital plane with integer coordinates $m, n$, is identified with a unit closed square and is called a pixel or cell. The pixels $p_1 = (m - 1, n)$, $p_2 = (m + 1, n)$, $p_3 = (m, n - 1)$ and $p_4 = (m, n + 1)$, are neighbours of the pixel $p = (m, n)$. Now let $V$ be a finite set of symbols, referred to as an alphabet and # a symbol not belonging to $V$. A two-dimensional picture array (or simply called an array) over $V$ is a finite set of connected pixels labeled by the symbols of $V$. It is sufficient for our purposes to give an array in pictorial form without mentioning the coordinates of the pixels. Note that by connected picture array we mean that every pixel labelled by a symbol of $V$ has at least another labelled pixel of the picture array as a neighbour. The symbol # is used as a label of a pixel to indicate that it is empty *i.e.* it is not occupied by any label of $V$. When specifying a picture array we can list the coordinates of the pixels of the array with their labels in a formal manner but it is enough to specify the labels of the pixels of the array. We denote the set of all non-empty connected finite picture arrays over $V$ by $V^{++}$. A subarray $Y$ of an array $X$ is a connected part of $X$. As an illustration, the picture array in Fig. 1($A$) with each pixel having label $a$, denotes a digitized form of the letter $H$ while in Fig. 1($B$), a labelled square array is shown with $d_1, d_2, f_1, f_2, p_v, d$ being the labels of the pixels. A "kolam" pattern can be thought of as a picture array in the two-dimensional plane as described in [5] with the cells of the picture array labelled by primitive "kolam" patterns [5], a list of which as shown in Fig. 2 and an example "kolam" is shown in Fig. 3. We note that based on the "Narasimhan's method" of "kolam" generation described in [8], a "kolam" pattern, which is composed of primitive "kolam" patterns, can be transformed into a picture array over the symbols of the primitive patterns and conversely. As an example, the picture array in Fig. 1($B$) corresponds to the "kolam" in Fig. 3. Nagata and Robinson [5] propose an interesting view of a "kolam" as being composed of primitive kolam patterns, a partial list of which is given in Fig. 2.

$$\begin{array}{cccccccc}
a &&&&&&& a \\
a &&&&&&& a \\
a &&&&&&& a \\
a & a & a & a & a & a & a & a \\
a &&&&&&& a \\
a &&&&&&& a \\
a &&&&&&& a
\end{array}$$

$(A)$

$$\begin{array}{ccccc}
d_1 & d_1 & d_1 & d_1 & d_1 \\
f_1 & d & d & d & f_2 \\
p_v & p_v & p_v & p_v & p_v \\
f_1 & d & d & d & f_2 \\
d_2 & d_2 & d_2 & d_2 & d_2
\end{array}$$

$(B)$

**Fig. 1.** $(A)$ An array representing the letter $H$; $(B)$ a square array

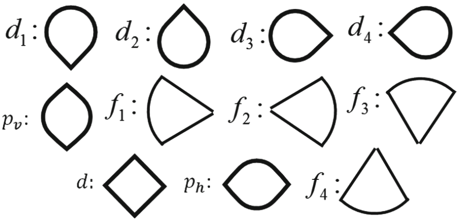


**Fig. 2.** Kolam primitive patterns.

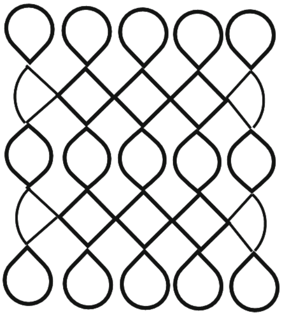


**Fig. 3.** A kolam pattern.

## 3 Contextual Array P System and "Kolam" Pattern Generation

In [3], contextual array P system is considered and several theoretical properties of this system are established. We first illustrate the manner of application of a rule of a contextual array grammar ($CAG$) [4] and then recall contextual array P system [3]. In a contextual array grammar ($CAG$) rule of the form $r : (\alpha, \beta)$, the arrays $\alpha$ and $\beta$ are referred to as "selector" and "context" respectively. We restrict ourselves to the case with the "selector" and the "context" in the contextual array grammar rule not having empty pixels *i.e.* the "selector" and the "context" are connected and labeled only by symbols from an alphabet $V$ and not by the blank symbol #. For arrays $X, Y \in V^{++}$ where $V$ is the alphabet, intuitively, if in $X$, we find a subarray that corresponds to the selector $\alpha$, and if the places (*i.e.* unit squares in the two-dimensional plane) corresponding to $\beta$ are labeled only by the blank symbol #, then we can add the context $\beta$ to $\alpha$, thus deriving an array $M$ and we write $X \Longrightarrow_G M$. If an array $Z$ is obtained from an array $X$, by a sequence of such steps, we write $X \Longrightarrow_G^* Z$. A maximal mode or $t$-mode of derivation, denoted $\Longrightarrow_G^t$ corresponds to collecting only the arrays produced by blocked derivations, namely, derivations which cannot be continued. For a formal definition of $CAG$ we refer to [4]. In the $*$-*mode* of derivation, all picture arrays derivable from an axiom are taken in the picture language (also called, array language) generated by $G$ while in the $t$-mode of derivation, all picture arrays produced by blocked derivations constitute the picture language. We give an example of a contextual array grammar working in $t$-mode generating the array language $L_1$ consisting of all picture arrays which are squares of side lengths $2n+3, n \geq 1$, with the label pixels of the picture arrays belonging to $V$. In such a picture array of $L_1$, the first row is of the form $d_1^{2n+3}$, the last row is of the form $d_2^{2n+3}$ and every two adjacent rows starting from the second row are of the form $f_1 d^{2n+1} f_2$ and $p_v^{2n+3}$ respectively. A $5 \times 5$ picture array of $L_1$ is shown in Fig. 1($B$) and represents the "kolam" in Fig. 3. We consider a contextual array grammar $G$ with an axiom picture array $A_1$, which is given in pictorial form: $A_1 := \begin{array}{ll} f_1 & \\ d_2 & d_2 \end{array}$ The contextual array rules of $G$ are given by the following pictorial forms where the pixels and symbols of the selector are indicated by enclosing them in boxes. In other words the rules $p_i (1 \leq i \leq 17)$ are given as follows:

$$p_1 := \begin{array}{ccc} & p & \\ \boxed{f_1} & d & d \\ \boxed{d_2} & \boxed{d_2} & \end{array}, p_2 := \begin{array}{ccc} & d & \\ \boxed{p} & p & p \\ \boxed{d} & \boxed{d} & \end{array}, p_3 := \begin{array}{ccc} & p & \\ \boxed{d} & d & d \\ \boxed{p} & \boxed{p} & \end{array}, p_4 := \begin{array}{ccc} & d_1 & \\ \boxed{\mathrm{d}} & d & f_2 \\ \boxed{p} & \boxed{p} & \end{array}, p_5 := \begin{array}{cc} \boxed{d_1} & d_1 \\ \boxed{d} & \boxed{f_2} \end{array}$$

$$p_6 := \begin{array}{cc} f_1 & \boxed{d} \\ \boxed{p} & \boxed{p} \\ \boxed{f_1} & \end{array}, \; p_7 := \begin{array}{cc} p & \boxed{p} \\ \boxed{f_1} & \boxed{d} \end{array}, \; p_8 := \begin{array}{cc} d & \boxed{d} \\ \boxed{p} & \boxed{p} \\ \boxed{d} & \end{array}, \; p_9 := \begin{array}{cc} p & \boxed{p} \\ \boxed{d} & \boxed{d} \end{array}, \; p_{10} := \begin{array}{cc} d_1 & \boxed{d_1} \\ \boxed{f_1} & \boxed{d} \end{array},$$

$$p_{11} := \begin{array}{cc} d_1 & \boxed{d_1} \\ \boxed{d} & \boxed{d} \end{array}, \; p_{12} := \begin{array}{cc} \boxed{d} & \boxed{d} \\ \boxed{d_2} & d_2 \end{array}, \; p_{13} := \begin{array}{cc} & \boxed{d} \\ \boxed{p} & \boxed{p} \\ \boxed{d} & d \end{array}, \; p_{14} := \begin{array}{cc} & \boxed{d} \\ \boxed{p} & \boxed{p} \\ \boxed{p} & p \end{array}, \; p_{15} := \begin{array}{cc} \boxed{d} & \boxed{f_2} \\ \boxed{d_2} & d_2 \end{array},$$

$$p_{16} := \begin{array}{cc} & \boxed{f_2} \\ \boxed{p} & \boxed{p} \\ \boxed{d} & f_2 \end{array}, \; p_{17} := \begin{array}{cc} \boxed{d} & \boxed{f_2} \\ \boxed{p} & p \end{array}$$

A maximal (that is, $t$-mode) derivation in $G$ generating a $5 \times 5$ picture array of the language $L_1$ is done by the application of the rules $p_1, p_2, p_4, p_5, p_7, p_8, p_6, p_{11}, p_{11}, p_{10}, p_{12}, f_{13}, p_{17}, p_{16}, p_{12}, p_{15}$ in this order. The first two steps of the derivation are shown below: $\begin{array}{ll} f_1 & \\ d_2 & d_2 \end{array} \Longrightarrow_{p_1} \begin{array}{lll} & p & \\ f_1 & d & d \\ d_2 & d_2 & \end{array} \Longrightarrow_{p_2}$

$\begin{array}{llll} & & d & \\ & p & p & p \\ f_1 & d & d & \\ d_2 & d_2 & & \end{array}$ It can be seen that the contextual array grammar rules of $G$ generate the picture language $L_1$. In an array $M$ of $L_1$ if we replace each of the symbols by the corresponding primitive "kolam" pattern as in Fig. 2, then we obtain the "kolam" itself. For example the rectangular array in Fig. 1($B$) yields the "kolam" in Fig. 3, when the symbols are substituted in this manner.

Informally expressed, the basic model of a rewriting array P system has a membrane structure $\mu$ of a P system with $m$ membranes, denoted by a well-formed expression of parentheses over the alphabet of left and right parentheses $[_i$ and $]_i$, $1 \leq i \leq m$. For example, $[_1\ [_2\ ]_2\ [_3\ ]_3\ ]_1$ means that membrane 1 is the outermost membrane (also called skin membrane), which contains membranes 2 and 3. The membranes (also called regions) have arrays as objects and rewriting rules as evolution rules, which can be applied to the objects. The application of a rule at the level of a membrane is sequential but the objects in all the membranes are rewritten in a maximal parallel way in the sense that all objects that can be rewritten in the membranes should be rewritten. There is a target indication *here*, *in* or *out* associated with every rule. The evolved array in a membrane is retained in the same membrane if the target is *here* and sent to an immediately inner membrane or outer membrane according as the target is *in* or *out*.

We now recall contextual array P system that has in its membranes, picture array objects and contextual array rules as in a contextual array grammar [4].

**Definition 1** [3]**.** *A contextual array P system with* $m \geq 1$ *membranes is a construct* $\Pi = (V, \#, \mu, X_1, \ldots, X_m, R_1, \ldots, R_m, i_o)$, *where* $V$ *is the alphabet;* $\#$ *is the blank symbol;* $\mu$ *is a membrane structure with* $m$ *membranes or regions, labeled by* $1, \ldots, m$, *in a one-to-one manner;* $X_1, \ldots, X_m$ *are finite sets of arrays over* $V$ *associated with the* $m$ *regions of* $\mu$*;* $R_1, \ldots, R_m$ *are finite sets of contextual array rules over* $V$ *associated with the* $m$ *regions of* $\mu$*; the rules have attached targets here, out, in, or* $in_j$, $1 \leq j \leq m$*;* $i_o$ *is the label of a membrane called output membrane, wherein the results of successful computations are collected.*

A *computation step* in a contextual array P system is done as follows: for each array $\mathcal{A}$ in each region of the system, if a contextual array production $p$ in the region can be applied to $\mathcal{A}$, then it should be applied which means that the application of a rule is sequential at the level of arrays, but maximally parallel at the level of the whole system. If more than one rule is applicable at the same time, then one is chosen in a nondeterministic way. The resulting array, if any, is placed in the region indicated by the target associated with the rule having been applied with interpreting the attached target as follows: *here* means that the array remains in the same region, *out* means that the array exits the current membrane and is placed in the immediately outer membrane if one exists (we do not allow the target *out* to be used by a rule assigned to the skin membrane), $in_j$ means that the array is immediately sent to the directly inner membrane with label $j$, and *in* means that the array is immediately sent to one of the directly inner membranes, chosen in a nondeterministic way if several such membranes exist (if no inner membrane exists, then a rule with the target indication *in* cannot be used). A computation is called *successful* if and only if it halts, which means that a configuration has been reached where no rule can be applied to the existing arrays. The result of a halting computation consists of the arrays collected in the membrane with label $i_o$ in the halting configuration.

A variant of contextual array P system is considered in [10]. In this variant, instead of a sequential application of a contextual array grammar rule, a set of

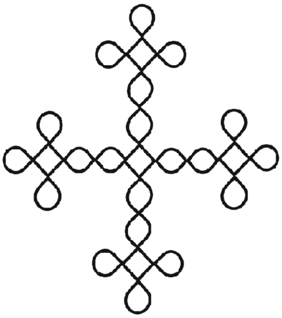


**Fig. 4.** P system “kolam”.

$$\begin{array}{ccccccccc}
 & & & & d_1 & & & & \\
 & & & d_3 & d & d_4 & & & \\
 & & & & p_v & & & & \\
 & d_1 & & & p_v & & & d_1 & \\
d_3 & d & p_h & p_h & d & p_h & p_h & d & d_4 \\
 & d_2 & & & p_v & & & d_2 & \\
 & & & & p_v & & & & \\
 & & & d_3 & d & d_4 & & & \\
 & & & & d_2 & & & &
\end{array}$$

**Fig. 5.** Array representing P system “kolam”.

rules in a membrane with the same target is applied in parallel. The resulting array P system is called Parallel contextual array P system ($PCAP$). It has been shown in [10], that the advantage of this kind of a $PCAP$ system is that the number of membranes required in the construction of such systems in generating picture languages, is reduced in comparison with the sequential $CAP$.

### 3.1 “Kolam” Generation Using *PCAP*

We now make use of $PCAP$ in generating “Kolam” patterns. We consider the kind of “kolam” shown in Fig. 4. This kind of a kolam can be represented by an array as shown in Fig. 5. We consider the picture language $L_2$ whose elements are picture arrays, one member of which is shown in Fig. 5. In fact in an array in $L_2$, the middle vertical column is a word of the form $d_1(dp_vp_v)^{n-1}d(p_vp_vd)^{n-1}d_2$, $n \geq 1$ but written vertically while every symbol $d$ in this column except the central symbol $d$ has the symbol $d_3$ to its immediate left and the symbol $d_4$ to its immediate right. Also, the middle horizontal row is a word of the form $d_3(dp_hp_h)^{n-1}d(p_hp_hd)^{n-1}d_4$, $n \geq 1$ while every symbol $d$ in this row except the central symbol $d$ has the symbol $d_1$ immediately above it and the symbol $d_2$ immediately below it. The picture language $L_2$ is generated by a $PCAP\ \Pi$ with a membrane structure $[_1\,[_2\,]_2\,]_1$ as in Fig. 6. The region labelled 1 has an axiom array $d$ while the region 2 has no array inside it initially. The region 2 is the output membrane where the picture arrays generated are collected which constitute the arrays in $L_2$. Region 1 has four rules $r_1, r_2, r_3, r_4$

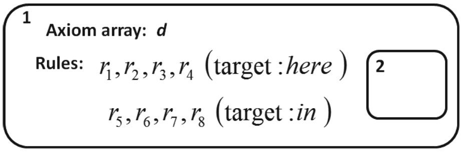


**Fig. 6.** Membrane structure

with target *here* and another four rules $r_5, r_6, r_7, r_8$ with target *in* while region 2 has no rules. The rules are given below:

$$r_1 := \begin{array}{ccc} d_3 & d & d_4 \\ & p_v & \\ & p_v & \\ & \boxed{d} & \end{array}, \; r_2 := \begin{array}{cccc} d_1 & & & \\ d & p_h & p_h & \boxed{d} \\ d_2 & & & \end{array}, \; r_3 := \begin{array}{ccc} & \boxed{d} & \\ & p_v & \\ & p_v & \\ d_3 & d & d_4 \end{array}, \; r_4 := \begin{array}{cccc} & & & d_1 \\ \boxed{d} & p_h & p_h & d \\ & & & d_2 \end{array}$$

$$r_5 := \begin{array}{c} d_1 \\ \boxed{d} \end{array}, \; r_6 := d_3 \, \boxed{d}, \; r_7 := \begin{array}{c} \boxed{d} \\ d_2 \end{array}, \; r_8 := \boxed{d} \, d_4$$

The generation of an array in $L_2$ as in Fig. 5 in the *PCAP Π* takes place as follows: Starting with the axiom array $d$ in region 1, the rules $r_1, r_2, r_3, r_4$ are applied in parallel generating the array

$$\begin{array}{ccccccc} & & d_3 & d & d_4 & & \\ & & & p_v & & & \\ d_1 & & & p_v & & & d_1 \\ d & p_h & p_h & d & p_h & p_h & d \\ d_2 & & & p_v & & & d_2 \\ & & & p_v & & & \\ & & d_3 & d & d_4 & & \end{array} .$$

The array remains in region 1 as the target indication of the rules $r_1, r_2, r_3, r_4$ is *here*. The process can be repeated. When the rules $r_5, r_6, r_7, r_8$ are applied in parallel, then the array generated is sent to region 2, due to the target *in* in the rules $r_5, r_6, r_7, r_8$. Since region 2 is the output membrane and does not contain any rule, the array is collected in the picture language $L_2$.

In every array in $L_2$ if we replace as in Narasimhan's method [8], the label (or symbol) of every pixel by the corresponding primitive "kolam" pattern from the list in Fig. 2, then we obtain a "kolam" which will be an enlarged version of the "kolam" in Fig. 4. In fact when the rules $r_1, r_2, r_3, r_4$ are applied in parallel once and the rules $r_5, r_6, r_7, r_8$ are then applied once in parallel, the array generated is as in Fig. 5. The "kolam" derived from this array by Narasimhan's method is indeed the "kolam" shown in Fig. 4.

We have shown that a class of "kolams", an example of which is shown in Fig. 4, can be generated by the P system, namely, parallel contextual array P system (*PCAP*) [10]. It can be shown that the same "kolam" class can be generated by a contextual array P system (*CAP*) [3] but it will require more number of membranes while the *PCAP* considered requires only two membranes and thus is a simpler model for "kolam" generation. But a *PCAP* with one membrane is not enough to generate this "kolam" class as the rules can have the only target *here* and so the rules $r_1$ to $r_8$ need not be applied in the combination of rules mentioned earlier, namely, $r_1, r_2, r_3, r_4$ together and $r_5, r_6, r_7, r_8$ together in parallel. But then the number of pixels in the vertical middle column in Fig. 4 above and below the central symbol $d$ as well as in the horizontal middle row in Fig. 4, in the right and left of the central $d$, need not be the same.

## 4 Concluding Remarks

We have considered the problem of generation of kolam-designs, in the framework of the computability model of P system. The intricate kolam designs are very good examples of visual expressions of creative thought [1] and can be considered as diagrams that could be classified under "art" in a very broad sense. The contextual array P ($CAP$) system model considered here makes use of certain kinds of rules of object generation, thus providing a novel approach to handle such kolam designs. It will be of interest to consider more complex "kolam" patterns and examine the ability of the contextual array P system in handling such patterns.

**Acknowledgement.** The authors thank the reviewers for their constructive comments and suggestions which very much helped to revise the paper and present the details in a better form. The first author Ibrahim Venkat acknowledges the support for this research by the RUI Grant Account # 1001/PKOMP/811290 awarded by Universiti Sains Malaysia. The third author K.G. Subramanian is grateful to University Grants Commission, India, for the award of Emeritus Fellowship (No. F.6-6/2016-17/EMERITUS-2015-17-GEN-5933 / (SA-II)) to him to execute his work in the Department of Mathematics, Madras Christian College.

## References

1. Ascher, M.: The kolam tradition. Am. Sci. **90**, 56–61 (2002)
2. Ceterchi, R., Mutyam, M., Paun, G., Subramanian, K.G.: Array-rewriting P systems. Nat. Comput. **2**(3), 229–249 (2003)
3. Fernau, H., Freund, R., Schmid, M.L., Subramanian, K.G., Wiederhold, P.: Contextual array grammars and array P systems. Ann. Math. Artif. Intell. **75**(1–2), 5–26 (2015)
4. Freund, R., Paun, G., Rozenberg, G.: Contextual array grammars. In: Subramanian, K.G., et al. (eds.) Formal Models, Languages and Applications, pp. 112–136. World Scientific Publishing, Singapore (2007)
5. Nagata, S., Robinson, T.: Digitalization of kolam patterns and tactile kolam tools. In: Subramanian, K.G., et al. (eds.) Formal Models, Languages and Applications, pp. 354–363. World Scientific Publishing, Singapore (2007)
6. Păun, G.: Membrane Computing: An Introduction. Springer, Heidelbrg (2000). https://doi.org/10.1007/978-3-642-56196-2
7. Rozenberg, G., Salomaa, A. (eds.): Handbook of Formal Languages, vol. 3. Springer, Berlin (1997). https://doi.org/10.1007/978-3-642-59126-6
8. Siromoney, G., Siromoney, R., Krithivasan, K.: Array grammars and kolam. Comput. Graph. Image Process. **3**(1), 63–82 (1974)
9. Subramanian, K.G., Saravanan, R., Robinson, T.: P systems for array generation and application to kolam patterns. Forma **22**, 47–54 (2007)
10. Subramanian, K.G., Bera, S., Song, B., Pan, L., Zhang, Z.: Array P systems based on parallel rewriting with array contextual rules. In: Pre-Proceedings of Asian Conference on Membrane Computing (ACMC 2017), pp. 403–415 (2017)

# Diagrams in Mathematics

# Visual Algebraic Proofs for Unknot Detection

Andrew Fish[1], Alexei Lisitsa[2], and Alexei Vernitski[3](✉)

[1] School of Computing, Engineering and Mathematics, Centre for Secure, Intelligent and Usable Systems, University of Brighton, Brighton, UK
andrew.fish@brighton.ac.uk
[2] Department of Computer Science, University of Liverpool, Liverpool, UK
A.Lisitsa@liverpool.ac.uk
[3] Department of Mathematical Sciences, University of Essex, Essex, UK
asvern@essex.ac.uk

**Abstract.** A knot diagram looks like a two-dimensional drawing of a knotted rubberband. Proving that a given knot diagram can be untangled (that is, is a trivial knot, called an unknot) is one of the most famous problems of knot theory. For a small knot diagram, one can try to find a sequence of untangling moves explicitly, but for a larger knot diagram producing such a proof is difficult, and the produced proofs are hard to inspect and understand. Advanced approaches use algebra, with an advantage that since the proofs are algebraic, a computer can be used to produce the proofs, and, therefore, a proof can be produced even for large knot diagrams. However, such produced proofs are not easy to read and, for larger diagrams, not likely to be human readable at all. We propose a new approach combining advantages of these: the proofs are algebraic and can be produced by a computer, whilst each part of the proof can be represented as a reasonably small knot-like diagram (a new representation as a labeled tangle diagram), which can be easily inspected by a human for the purposes of checking the proof and finding out interesting facts about the knot diagram.

## 1 Introduction

A knot diagram looks like a two-dimensional drawing of a knotted rubberband. In the simplest case, consider a knot diagram (or a rubberband) without crossings (e.g. see the lower right diagram in Fig. 1); this knot diagram is known as the *trivial knot* or *unknot*. For this and other basic concepts of knot theory, see any of the textbooks [1,9,10,12,13,16]. A knot diagram which looks knotted may be really knotted, or it may be then the unknot in disguise, with the diagram (think rubberband) being able to be untangled by gently pulling some of its parts in some order, until the diagram does not have any crossings (at which point it is obvious that it is a trivial knot). An example of a sequence of such untangling steps is shown in Fig. 1; intuitively these correspond to moving a rubberband in the 3-dimensional Euclidean space without cutting it.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 89–104, 2018.
https://doi.org/10.1007/978-3-319-91376-6_12

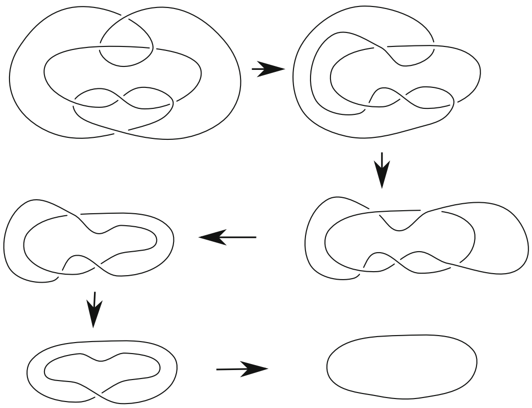


**Fig. 1.** An example of a step-by-step untangling of a diagram of a trivial knot.

The problem of deciding whether a given knot diagram can be untangled (that is, is a trivial knot) is one of the most famous problems of knot theory. It is an interesting problem, having attracted diverse approaches from a number of different areas of mathematics, whilst having an immediate aesthetic appeal.

The immediate, naive approach to the problem is trying to find a sequence of untangling moves explicitly; then the process of untangling can be represented as a sequence of diagrams (as in Fig. 1). For small knot diagrams this can be a preferred method, but for larger knot diagrams producing such proofs is difficult [2], and the produced proof becomes hard to inspect (as you can see, even the relatively small proof in Fig. 1 has some steps that are not so easy to follow).

A number of more advanced approaches are based on using algebra. Some of these approaches are based on denoting each arc (that is, a continuous unbroken line segment of the diagram) by a letter, as shown in the example in Fig. 2, and then proving that all these letters are equal to each other in a certain algebra. Some of the authors' previous research concentrated on such methods of untangling [4,5,11], and this paper presents a new twist in this research. (For completeness, note that not all algebraic approaches to untangling are based on labelling arcs; see [3], for example.)

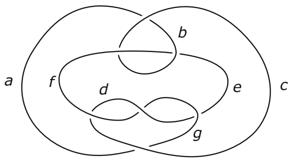


**Fig. 2.** A labelled knot diagram

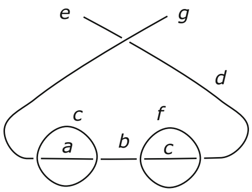


**Fig. 3.** A labelled tangle diagram

The advantage of such an approach is that since proofs are purely algebraic, a computer can be used to produce the proofs, and, therefore, a proof can be produced even for a large knot diagram. However, the produced proof, consisting of long chains of abstract equalities, is not easy to read and, for larger diagrams, is not human readable at all. For example, here is a proof showing that the diagram in Fig. 2 is a trivial knot (adapted from [19]):

"$bc = ca = gc$, hence, $b = g$; $bf = fc = df$, hence, $b = d$; $gg = gb = gd = eg$, hence, $g = e$; $dg = fd = fb = fg$, hence, $d = f$; $cf = fd = dg = db = dd = df$, hence, $c = d$; $ab = ad = ac = cb$, hence, $a = c$."

In this paper, we propose a new approach, which combines the advantages of the above approaches. Our proofs are produced algebraically (and can be produced by a computer), as in the example above, but they are produced in such a way that the proof of the equality of each two letters (say, $a$ and $b$) can be represented as a reasonably small knot-like diagram (with two free ends, labelled $a$ and $b$), which can be easily inspected by a human for the purposes of checking the proof and finding out interesting facts about the knot diagram. For instance, instead of reading the proofs of $b = g$, $b = d$ and $g = e$ above, one can inspect the diagram in Fig. 3; we shall revisit this example in Sect. 4.

The following sections explain how such labelled tangle diagrams are related to the knot diagram, and why the existence of certain labelled tangle diagrams proves that a knot diagram represents the trivial knot.

Since this research involves considering many types of diagrams, it may be useful to highlight how different diagrams are used for different purposes in this paper.

- There are knot diagrams, such as in Fig. 2.
- There are tangle diagrams with two free ends, such as in Fig. 3; the existence of certain tangle diagrams with two free ends (such as, for example, in Figs. 11, 12, 13, 14, 15 and 16) proves that the knot diagram is the trivial knot.
- We have proved Theorem 3 which establishes a connection between knot diagrams and tangle diagrams by considering certain more complicated diagrams, which look like tangle diagrams with one special strand which we call a virtual ruler; see Sect. 3.
- In practice, finding suitable tangle diagrams with two free ends involves considering other tangle diagrams (not only with two free ends), as described in Sects. 5 and 6.

## 2 Groups Induced by a Knot Diagram

This section presents definitions, together with an expanded and corrected exposition of Fact 1 and the subsequent discussion in [11].

By an *arc* we mean a continuous line in a knot diagram from one undercrossing to another undercrossing. For example, consider the knot diagram in Fig. 2; it has seven arcs, denoted by $a, b, \ldots, g$.

For a given knot diagram $D$, the $\pi$-orbifold group $OD$ of the knot is a group generated by the arc letters with the following relations. For each arc $x$ of the diagram $D$, introduce a relation $x^2 = 1$. At every crossing where $x$ and $z$ are the two arcs terminating at the crossing and $y$ is the arc passing over the crossing, introduce a defining relation $xy = yz$ (or, equivalently, $yx = zy$, or $yxy = z$, or $yzy = x$). Let $A$ denote the generating set of $OD$ (i.e. the set of labels of the arcs of $D$), and consider the natural homomorphism from the free semigroup $A^+$ onto $OD$. It is easy to see that, for each element $g$ of $OD$, either only words of an odd length are mapped to $g$ or only words of an even length are mapped to $g$ under the homomorphism. Accordingly, let us say that $g$ is an element of odd (even) length in the former (latter) case. A subgroup of $OD$ consisting of the set of all elements of even length is called the fundamental group of the 2-fold branched cyclic cover space of a knot [15,20]; we shorten this name to the *two-fold group of a knot*, and denote the group by $TD$.

Another well-known algebraic construction associated with a knot diagram $d$ is its *knot group*, which we denote by $GD$. This is historically the first and the best known construction (see, for example, Sect. 6.11 in [6] or Chap. 11 in [10]). We do not need to define $GD$ here, but we note that $OD$ is a factor-group of $GD$ produced from $GD$ by introducing the additional relations $x^2 = 1$ for each arc $x$.

Trivial knots can be characterised via certain algebraic constructions associated with them.

**Theorem 1.** *The following are equivalent:*

- *A knot diagram $D$ is a diagram of the trivial knot.*
- *The two-fold group $TD$ is trivial* [15,20]*. In other words, for each pair of arc labels $x$ and $y$, we have $x = y$ in the two-fold group of the knot.*
- *The group $GD$ is infinite cyclic* [1]*. In other words, for each pair of arc labels $x$ and $y$, we have $x = y$ in the knot group.*

Since the $\pi$-orbifold group $OD$ is 'sandwiched' between the two-fold group of the knot $TD$ and the group of the knot $GD$, the following conclusion can be made concerning $\pi$-orbifold groups.

**Corollary 1.** *A knot diagram $D$ is a diagram of the trivial knot if and only if its $\pi$-orbifold group $OD$ is the two-element cyclic group. In other words, $D$ is a diagram of the trivial knot if and only if for each pair of arc labels $x$ and $y$, we have $x = y$ in $OD$.*

The rest of the theoretical discussion in the paper concentrates on discussing how one can prove, for any two given arc labels $x$ and $y$, that $x = y$ in the $\pi$-orbifold group of the knot.

## 3 Reading Tangles

Tangle diagrams have been used for untangling knot diagrams; see, for example, [8]; in this paper we introduce a completely new way of using tangles to untangle knot diagrams. In this section, we develop the theory of reading tangles. This takes the form of flexible rulers, called virtual rulers, each of which gives rise to a word in the $\pi$-orbifold group. Then, the passing of such a ruler through a tangle

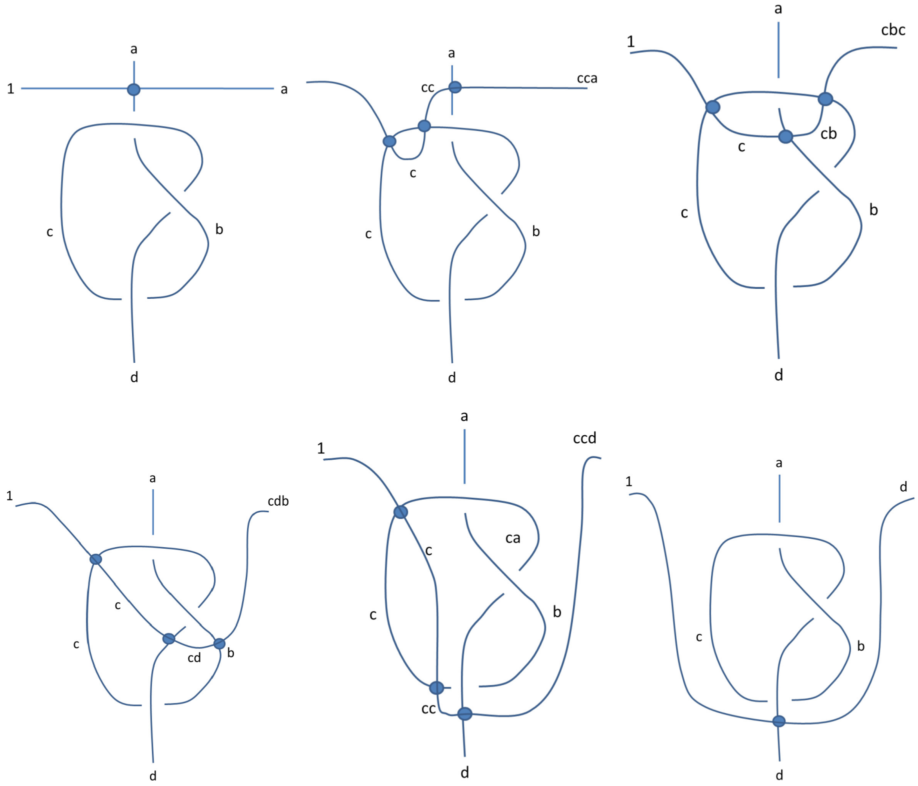


**Fig. 4.** An example of passing a virtual ruler over a tangle diagram to produce a proof. Reading the words induced by the rulers starting in the top row from left to right, followed by the bottom row gives rise to the proof: $a = cca = cbc = cdb = ccd = d$.

diagram corresponds to a proof in the group. We first provide some examples to illustrate the core idea, before moving into the theoretical developments ensuring that the intuitive use of these tangle diagrams and virtual rulers is well-founded.

Figure 4 shows an example of passing a virtual ruler (the line across the tangle, from left to right, with meeting points of the ruler and the diagram accentuated with dots) over a tangle diagram to produce a proof. The tangle diagram has its arcs labeled, and the labels on the ruler concatenate the labels encountered as it passes through the tangle, starting with 1 on the left hand side of the ruler playing the role of the empty word. The tangle diagrams in all six cases in Fig. 4 are the same and the rulers differ in a specific way (details to follow). We obtain a proof consisting of equalities of the words from each of the six cases. In this example, reading the words induced by the rulers starting in the top row from left to right, followed by the bottom row gives rise to the proof: $a = cca = cbc = cdb = ccd = d$.

Later on, we will see that each step between diagrams can be viewed precisely as a certain type of move, called a $TR_2$ or a $TR_3$ move (these will be defined soon), and we will use this characterization to ensure the equalities we claim in the proof are correct. Alternate choices of application of the $TR_2$ and $TR_3$ moves can give rise to different proofs, as demonstrated in Fig. 5. The first four diagrams and rulers in the sequences are the same in Figs. 4 and 5, but the fifth is different, effectively by applying a different $TR_3$ move to the fourth diagram in the sequence. From Fig. 5, we obtain the slightly different proof: $a = cca = cbc = cdb = dbb = d$.

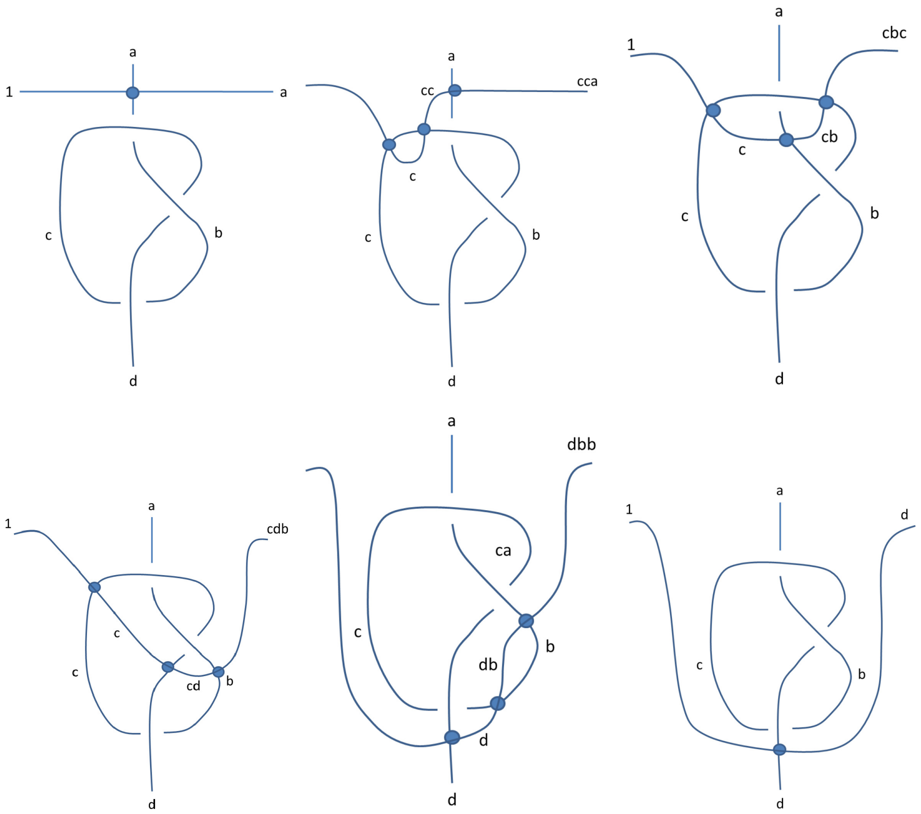


**Fig. 5.** A different proof obtained by applying a different $TR_3$ move to the 4th diagram in the sequence in Fig. 4. We obtain the proof: $a = cca = cbc = cdb = dbb = d$.

### 3.1 The Theory of Reading Tangles

A *tangle* diagram $T$ is like a knot diagram, except that its arcs may have free ends (commonly arranged at the top and bottom of a bounding box of the diagram –

this is not essential, as can be seen from some examples in the rest of the paper, but we adopt this convention here to make the exposition more straightforward). The arcs are labeled by elements of the group. For our construction, we require that each labeled crossing matches exactly with one of the labeled crossings of the original knot diagram.

Figure 6 shows an example of a tangle diagram with one free end at the top and one at the bottom. This is the tangle considered in Figs. 4 and 5. In this paper, we only need to consider tangles with one free end at top and bottom, but the concept generalizes, as does the theory of reading tangles developed here.

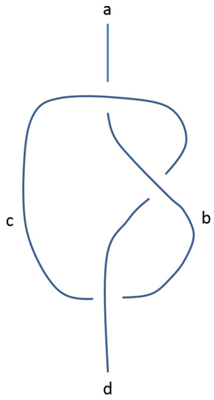


**Fig. 6.** The tangle diagram which is used in all cases of the Figs. 4 and 5 – the additional part of those figures that changed was the virtual ruler.

**Definition 1.** *A **virtual ruler** $v$ for $T$ is an additional strand with one free end at the left and one at the right of $T$, which only meets $T$ transversely at points that are not classical crossings.*

This means that the ruler can be viewed as a line drawn through a tangle diagram from the left to the right of the tangle, which crosses the tangle diagram properly (so no tangential meetings or concurrency of line segments) and does not pass through any already existing classical crossing of the tangle diagram. The meeting points between tangle diagram and the ruler are clearly indicated via blobs (which may be called virtual crossings).

**Definition 2.** *Let $T$ be tangle diagram and $v$ a virtual ruler for $T$. Then the arcs of $v$ have labels in the group induced from $T$ by assigning the identity (denoted by 1 here) to the leftmost free arc of $v$, and concatenating the labels on the arcs of $T$ that $v$ crosses as one traverses the virtual ruler from left to right. The word $w$ obtained as the label on the rightmost free end of $v$ is the **interpretation** of $v$ in $T$.*

The arcs of $v$ end at the blobs. Here, since we are considering a group we made use of the identity as the leftmost label of the ruler (whilst the empty word $\epsilon$ can be used here for the more general setting).

In Fig. 7, we present moves of virtual rulers over tangles (the definition follows). These will be precisely the moves of virtual rulers and tangles that encapsulate equivalence (see Theorem 2).

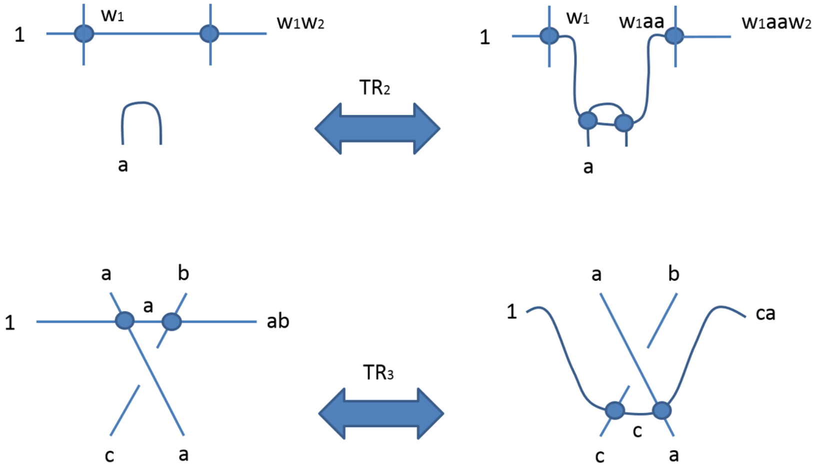


**Fig. 7.** The TR-moves, indicating permissible moves of a virtual ruler over a tangle.

**Definition 3.** *Let $T$ be tangle diagram and $v$ a virtual ruler for $T$. Define two moves of a virtual ruler over $T$, denoted $TR_2$ and $TR_3$, as shown in Fig. 7. For $TR_2$, the $w_1$ and $w_2$ are words on the labels of the arcs of the virtual ruler shown; the case in which either (or both) of $w_1$ and $w_2$ are empty is also permitted (if the left or right virtual crossing shown, respectively, is not present). For $TR_3$, the analogous move with a different classical crossing, shown in Fig. 8 is also permitted.*

**Theorem 2.** *Given any two virtual rulers $v_1$ and $v_2$ for $T$, the interpretation of $v_1$ in $T$ and the interpretation of $v_2$ in $T$ are equal in $OD$.*

*Proof.* In the same way that one can pass a strand of a knot diagram over the rest of the diagram by repeated application of Reidemeister moves $R_2$ and $R_3$, one can pass a virtual ruler over $T$ by the repeated application of the $TR_2$ and $TR_3$ moves. We see that each of these moves induces an equality in $OD$ by precisely the application of one of the defining relations. For $TR_2$, it is the application of the relation $a^2 = 1$, with $a$ a generator. For $TR_3$ it is the relation $ab = ca$, obtained from the labeled crossing in $T$, which by construction occurred in $OD$ (see the matching crossing which is the second on the left in Fig. 10).

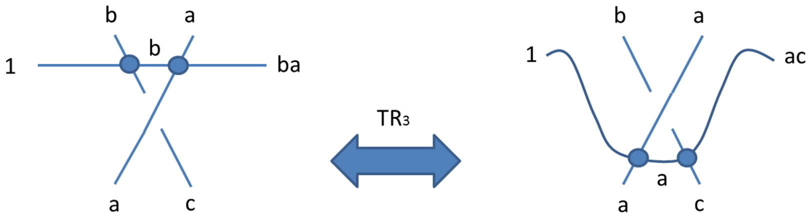


**Fig. 8.** An analogous $T_3$ move (we do not distinguish in naming convention).

**Lemma 1.** *The rules* $TR_2$ *and* $TR_3$ *are sufficient to pass a virtual ruler over a tangle.*

*Proof.* Consider the stepwise process of moving a virtual ruler over a fixed tangle by moving *segments* down the tangle from the top to the bottom. By *segment* here we refer to a connected part of the virtual ruler, not necessarily starting or ending at virtual crossings, which contains at most two virtual crossings. Any such segment considered contains 0, 1 or 2 virtual crossings. In the case of 0 virtual crossings, only $TR_2$ can be applied (see top row of Fig. 9). In the case of 2 virtual crossings, either we are in the case where $TR_3$ can be directly applied (if there is a $\Delta$ with the two virtual crossings and a classical crossing at the corners and there are no other strands meeting the $\Delta$, as in the middle row of Fig. 9), or not. If not, then consider a segment containing only one of the strands meeting the virtual ruler at one of these two virtual crossings, reducing to the case of 1 virtual crossing. In this case, apply a nearby $TR_2$ move, which enables the subsequent application of a $TR_3$-move (see bottom row of Fig. 9). Any $TR_3$-move applied reduces the number of classical crossings remaining to pass over. The process terminates.

Recall that a *proof* in $OD$ is a sequence of equalities of elements of $OD$ such that each consecutive element differs by the application of one of the relators in $OD$.

**Proposition 1.** *Let* $T$ *be a tangle diagram, and let* $v_1, \ldots, v_k$ *be a sequence of virtual rulers for* $T$*, with interpretations* $i(v_1), \ldots, i(v_k)$*. Then:*

1. *The sequence of equalities* $i(v_1) = \ldots = i(v_k)$ *holds in* $OD$*.*
2. *If, in addition, for each* $j \in \{1, \ldots, k-1\}$ *we have that* $v_j$ *differs from* $v_{j+1}$ *by a* $TR$ *move (either* $TR_2$ *or* $TR_3$*), then the sequence of equalities* $i(v_1) = \ldots = i(v_k)$ *is a proof in* $OD$*.*

*Proof.* The first part follows from Theorem 2, whilst the second part follows from the observation in the proof of Theorem 2 about the matching of the $TR$ moves with the relators.

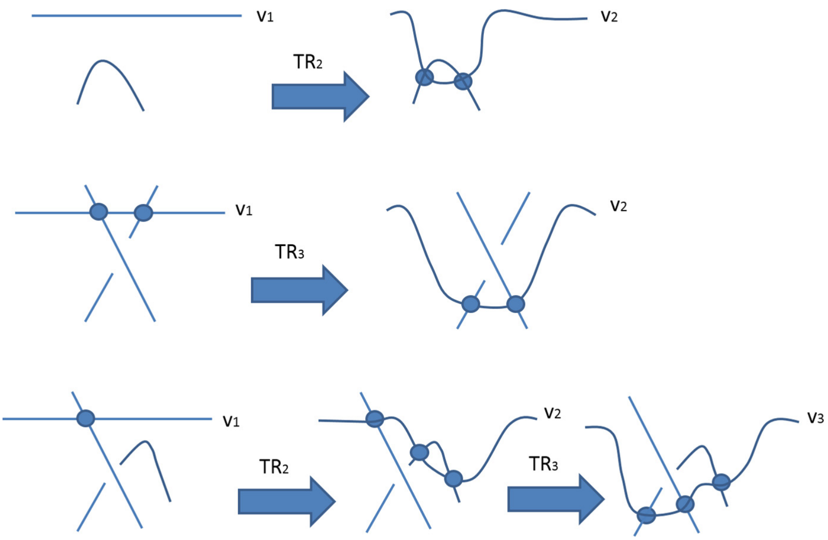


**Fig. 9.** The process of moving a virtual ruler over a tangle.

**Corollary 2.** *There exist different proofs obtained from one fixed $T$ corresponding to different sequences of virtual rulers over $T$.*

*Proof.* See Figs. 4 and 5.

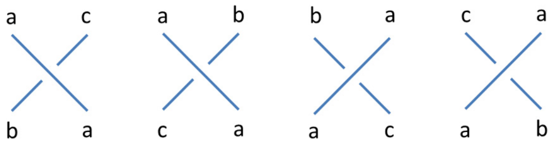


**Fig. 10.** Relations, from left to right: $ac = ba, ab = ca, ba = ac, ca = ab$.

## 4 Untangling: The Main Result and an Example

The theory we developed gives us a practical method for proving that a knot diagram is a diagram of the trivial knot by drawing certain tangle diagrams.

**Theorem 3.** *A knot diagram $D$ (with unique labels) represents the trivial knot if and only if for each pair of its labels $a, b$ there is a labelled tangle diagram $T$ which has exactly two free-end arcs labelled $a$ and $b$, with the property that each crossing in $T$ is labelled in the same way as some crossing in $D$.*

*Proof.* The 'if' direction follows directly from Corollary 1 and Theorem 2. The 'only if' direction follows from the fact that every derivation of an equality of two letters $x = y$ in $OD$ naturally induces a tangle diagram with two free ends $x$ and $y$, with relations of $OD$ being transformed into crossings as in Fig. 10.

We present a small example, produced manually. In the next section we report on our progress with building tangle diagrams using the computer.

*Example 1.* The knot diagram in Fig. 2 represents the trivial knot.

The proof splits into several lemmas, each demonstrating that two arc labels (that is, two generators of $OD$) are equal. The lemmas and their proofs are presented in Figs. 11, 12, 13, 14, 15 and 16; note that these diagrams are not illustrations of proofs, but actual proofs: that is, the existence of a diagram presented in Fig. 11 is, according to Theorem 3, a proof that $b = g$, and so on. For brevity of presentation, in some of the lemmas, a rectangle marked $i$ is used as shorthand, meaning that the diagram from Lemma $i$ should be substituted for this rectangle. This convention enables us to make diagrams more compact. As an example of the use of rectangles, compare the diagram in the lemma in Fig. 13 with the equivalent diagram in Fig. 3, which presents the same step in the proof. Whilst the equalities are immediate from the earlier results, the interested reader can also directly compare the equalities read off from each tangle diagram as one traverses from one free end to the other (e.g. from $b$ to $g$ in Fig. 13) and compare with the algebraic proof presented earlier (e.g. "$bc = ca = gc$, hence, $b = g$").

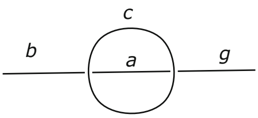


**Fig. 11.** $b = g$

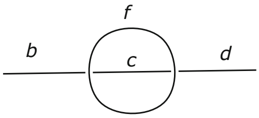


**Fig. 12.** $b = d$

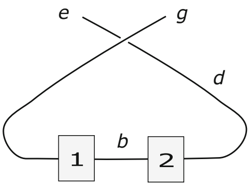


**Fig. 13.** $e = g$

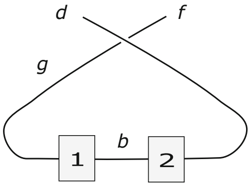


**Fig. 14.** $d = f$

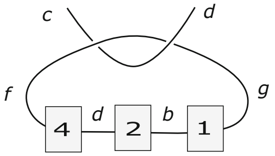


**Fig. 15.** $c = d$

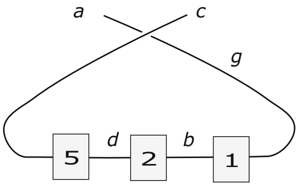
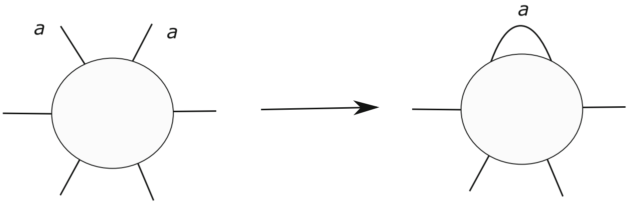


**Fig. 16.** $a = c$

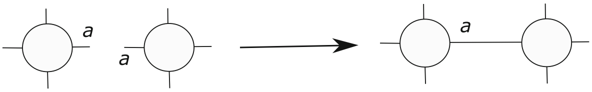
## 5 Manipulating Tangle Diagrams with the Computer

One can observe that the tangle diagrams in proofs (for example, like those in Fig. 3) may be assembled from copies of individual crossings of the original knot diagram by applying steps of the following two types:

- Given a labelled tangle diagram which has, among its end arcs, two adjacent end arcs labelled with the same letter, connect these two arcs.



- Given two labelled tangle diagrams $T$ and $U$ such that both $T$ and $U$ have an end arc labelled with the same letter, connect these two arcs.



To do this using the computer, instead of tangle diagrams themselves, we consider words which one can read on end arcs around a tangle diagram (clockwise or anticlockwise, starting from any point outside the diagram). To start with, we list all words which one can read around individual crossings of the knot diagram such as, for example, *acbc* around the top crossing of the diagram on Fig. 2 (for the top crossing walking in a small circle around the crossing meet four arcs and the labels are recorded in the order the arcs are met). Whenever we have a word $a_1 a_2 \dots a_{n-1} a_n$, we also produce the word $a_2 \dots a_{n-1} a_n a_1$ (which is a cyclic shift of the original word, corresponding graphically to starting reading from a different point outside the diagram) and the word $a_n a_{n-1} \dots a_2 a_1$ (which is the original word inverted, corresponding graphically to reading around the tangle in the opposite direction)[1]. In addition to these two 'trivial' ways of producing new words, we have two more, corresponding to the graphical moves above. Whenever we have a word $a_1 a_2 a_3 \dots a_n$ and $a_1 = a_2$, we produce the word $a_3 \dots a_n$ (which corresponds to connecting two adjacent arcs with identical labels). Whenever we have two words $a_1 a_2 \dots a_n$ and $b_1 b_2 \dots b_n$ and $a_1 = b_1$, we produce the word $a_2 \dots a_n b_2 \dots b_n$ (which corresponds to connecting two tangle diagrams)[2].

---

[1] In the implementation presented in Sect. 6 we did not use word reversion because, although reversing a word may make a proof shorter, it is not easy to implement it in the prover software we used.

[2] Another approach that we tested (as presented in Sect. 6) is simply to place two tangle diagrams next to each other without connecting them by an arc; this corresponds, in terms of labelling words, to concatenating the words.

The aim of this process is to produce all (or, to be more precise, 'sufficiently many') two-letter words $ab$. In our computer experiments (as presented in Sect. 6), we enumerated all arc labels of the knot diagram in some order and proved each equality $a_i = a_{i+1}$ by proving that we can produce the word $a_i a_{i+1}$.

## 6 Automated Proofs

We experimented with the procedure described in Sect. 5 by using its reduction to automated theorem proving. To a given knot diagram $D$ we can associate a first-order theory $T_D$ in a vocabulary which consists of unary predicate symbol $T$, binary functional symbol $*$ and constants $e$ and $a_1, \ldots, a_k$ for all of the labels of $D$. The axioms of $T_D$ include:

**I.** Axioms of monoid for $(*, e)$:
  - $(x * y) * z = x * (y * z)$ (associativity of multiplication)
  - $x * e = e * x = x$ ($e$ is a unit of the monoid)

**II.** – $a_i^2 = e$ for all labels $a_i$

**III.** Initial state axioms:
  - $T(a_i * a_j * a_k * a_j)$ for all crossings in $D$, where the over-crossing arc is labelled by $a_j$ and the under-crossing arcs are labelled by $a_i$ and $a_k$.

**IV.** Transition axioms:
  - $T(a_i * x) \rightarrow T(x * a_i)$ for all arc labels $a_i$
  - $T(x)$ & $T(y) \rightarrow T(x * y)$.

The ground terms are built from constants by the monoid operation, and are meant to represent words read on the ends of arcs around the tangle diagrams. The intended meaning of $T(w)$ for a word $w$ is a corresponding diagram (i.e. with $w$ read on its end arcs) which can be built following the rules in Sect. 5. The initial state axioms declare that the original crossings present us with the initial building blocks of the tangle diagrams to be able to start construction. The transition axioms describe permissible operations for building new tangles. The following result holds.

**Proposition 2.** *A knot diagram $D$ with arcs labelled by $a_1, \ldots, a_k$ is a diagram of trivial knot (unknot) if and only if $T_D \vdash \bigwedge_{1 \leq i \leq k-1} T(a_i * a_{i+1})$, where $\vdash$ denotes first-order logic derivability.*

*Proof.* Due to Theorem 3 it is sufficient to show that $T_D \vdash T(a_i * a_{i+1})$ iff a labelled tangle diagram can be built for $D$ with two free end arcs labelled by $a_i$ and $a_{i+1}$ using the procedure from Sect. 5.

$\Leftarrow$: Assume that a tangle diagram with free ends labelled by a word $w$ is built by the procedure. Then straightforward induction on the length of construction shows that $T_D \vdash T(t_w)$, where $t_w$ is a term encoding of $w$, that is a term built from $a_1, \ldots, a_k$ using $*$. Indeed, the statement holds for initial tangle diagrams formed by original crossings due to axioms **III**, which are included in $T_D$. For the inductive step, we assume that the statement holds for two tangle diagrams labelled by words $w_1$ and $w_2$, so we have that $T_D \vdash T(t_{w_1})$ and $T_D \vdash T(t_{w_2})$. Then, if $w'$ labels a tangle diagram obtained by connecting two adjacent arcs in the tangle labelled by $w_1$ we have that $T_D \vdash T(t_{w'})$, using the induction hypothesis for $w_1$ and the axioms **II** and **I**. If a tangle diagram labelled by $w''$ is obtained by connecting tangle diagrams labelled by $w_1$ and $w_2$ then $T_D \vdash T(t_{w''})$ using the inductive hypothesis for $w_1$ and $w_2$ and the transition axioms **IV**.

$\Rightarrow$: By induction on the number of applications of *Modus Ponens* rule using a transition axiom $T(x)\&T(y) \rightarrow T(x * y)$ we show that if $T_D \vdash T(t_w)$ then a tangle diagram can be constructed by the procedure from Sect. 5, which is labelled by $w'$ such that **I, II** $\vdash t_w = t_{w'}$. Indeed, for the base of induction we notice that the only formulae of the form $T(\ldots)$ derivable form $T_D$ without applying *Modus Ponens* to $T(x)\&T(y) \rightarrow T(x * y)$ are formulae $T(t_{w'})$ with $t_{w'}$ such that **I, II** $\vdash t_w = t_{w'}$ for some $t_w$ from an initial state axiom $T(t_w)$. Then the crossing corresponding to this axiom provides with the required tangle.

For the step of induction consider a derivation $T(t_w * t_{w'})$ from already derived $T(t_w)$ and $T(t_{w'})$ and the transition axiom $T(x)\&T(y) \rightarrow T(x * y)$. By the induction assumption we have required tangle diagrams constructed for $T(t_w)$ and $T(t_{w'})$, that is diagrams labelled by $w$ and $w'$, respectively. Then the tangle diagram for $T(t_w * t_{w'})$ is constructed by placing latter diagrams next to each other, and it is labelled by $ww'$. Further derivations of formulae of the form $T(..)$ from $T(t_w * t_{w'})$ using axioms **I, II** and first transition axiom are possible. In all such cases the required tangle diagram will be either the same as for $T(t_w * t_{w'})$, or reduced by connecting two adjacent arcs with equal labels.

Proposition 2 suggests a procedure for establishing unknottedness by using automated provers for first-order logic. Given a knot diagram $D$, specify a theory $T_D$ and apply an automated theorem prover to $T_D \vdash \bigwedge_{1 \leq i \leq k-1} T(a_i * a_{i+1})$.

In Table 1 we report on experiments[3] using the automated prover Prover9 [14] on a few well-known unknot diagrams.

The authors have used the same software to implement other algebraic techniques (similar to the one described in Corollary 1) for proving unknottedness of knot diagrams [4,11]. For the diagrams listed in Table 1, other techniques prove unknottedness faster (as expected). However (also as expected), the proofs of unknottedness produced by the methods proposed in this paper are more transparent and more amenable to the interpretation as untangling sequences.

---

[3] System used in experiments: Intel(R) Core(TM) i7-4790 CPU 3.60 Ghz, RAM 32 GB, Windows 7 Enterprise.

**Table 1.** Time taken to prove unknottedness of some known diagrams

| Name of unknot | Reference | # of crossings | Time, s |
|---|---|---|---|
| "Trivial" Trefoil | [8] | 3 | 0.01 |
| No name | Fig. 2 | 7 | 0.09 |
| Culprit | [8] | 10 | 5.16 |
| Goerlitz | [7] | 11 | 6.38 |
| Thistlethwaite | [18] | 15 | 321.0 |
| Ochiai, I | [17] | 16 | 1286.1 |

## 7 Conclusion and Future Work

In the paper, we proposed a novel use of labelled tangle diagrams as a means of representing certain algebraic proofs. This provides the opportunity for providing a visual readable proof instead of an algebraic textual proof, such that the proof notation is of a similar type to the original type of diagrams considered. We developed theory to demonstrate the equivalence of the use of these tangle diagrams to express correct algebraic equalities, in effect showing that proofs in the algebra can be represented visually via these labelled tangles. This was performed making use of some new, reusable machinery (virtual rulers, $TR$-moves) that will have independent uses. Whilst such proofs for small scale diagrams can be manually developed (as per the example in Sect. 4), there is a definite need for computer assistance on the larger scale. To this end, we present progress in automating the production of such proofs. We provide an indication of the time taken for the automate search for such proofs, and we observe a likely trade-off in the form of a slow-down of unknot detection in order to be able to develop the new visual proofs rather than algebraic proofs only.

This work opens us several avenues for future research. Firstly, extracting readable proofs from the output of a general-purpose prover software is very difficult. To help with this task, after the prover software has produced proofs, another script is needed to convert the proofs into a readable form. Alternatively, we can write our own specialist prover code. Secondly, producing nice-looking tangle diagrams from proofs manually is time consuming. For larger diagrams, the actual diagrams must be produced by the computer. After such tool support is present, an exploration of the utility of the representation and the tool by different possible user groups (e.g. undergraduate students versus research mathematicians) will also become feasible.

## References

1. Burde, G., Heusener, M., Zieschang, H.: Knots. De Gruyter (2013)
2. Coward, A., Lackenby, M.: An upper bound on reidemeister moves. Am. J. Math. **136**(4), 1023–1066 (2014)

3. Dynnikov, I.A.: Three-page approach to knot theory. Universal semigroup. Funct. Anal. Appl. **34**(1), 24–32 (2000)
4. Fish, A., Lisitsa, A.: Detecting unknots via equational reasoning, I: exploration. In: Watt, S.M., Davenport, J.H., Sexton, A.P., Sojka, P., Urban, J. (eds.) CICM 2014. LNCS (LNAI), vol. 8543, pp. 76–91. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-08434-3_7
5. Fish, A., Lisitsa, A., Stanovský, D.: A combinatorial approach to knot recognition. In: Horne, R. (ed.) EGC 2015. CCIS, vol. 514, pp. 64–78. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25043-4_7
6. Gilbert, N.D., Porter, T.: Knots and Surfaces. Oxford University Press, Oxford (1994)
7. Kauffman, L.H., Henrich, A.: Unknotting unknots. https://arxiv.org/abs/1006.4176
8. Kauffman, L.H., Lambropoulou, S.: Hard unknots and collapsing tangles. https://arxiv.org/abs/math/0601525v5
9. Kawauchi, A.: A Survey of Knot Theory. Birkhäuser, Basel (1996). https://doi.org/10.1007/978-3-0348-9227-8
10. Lickorish, W.B.R.: An Introduction to Knot Theory, vol. 175. Springer Science & Business Media, New York (2012). https://doi.org/10.1007/978-1-4612-0691-0
11. Lisitsa, A., Vernitski, A.: Automated reasoning for knot semigroups and $\pi$-orbifold groups of knots. In: Blömer, J., Kotsireas, I.S., Kutsia, T., Simos, D.E. (eds.) MACIS 2017. LNCS, vol. 10693, pp. 3–18. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-72453-9_1
12. Livingston, C.: Knot Theory, vol. 24. Cambridge University Press, Cambridge (1993)
13. Manturov, V.: Knot Theory. CRC Press, Boca Raton (2004)
14. McCune, W.: Prover9 and Mace4 (2005–2010). http://www.cs.unm.edu/~mccune/prover9/
15. Morgan, J.W., Bass, H. (eds.): The Smith Conjecture. Elsevier, New York (1984)
16. Murasugi, K.: Knot Theory and Its Applications. Springer Science & Business Media, New York (2007). https://doi.org/10.1007/978-0-8176-4719-3
17. Ochiai, M.: Non-trivial projections of the trivial knot. http://repository.kulib.kyoto-u.ac.jp/dspace/handle/2433/99940
18. Unknot. https://en.wikipedia.org/wiki/Unknot
19. Vernitski, A.: Describing semigroups with defining relations of the form $xy = yz$ and $yx = zy$ and connections with knot theory. Semigroup Forum **95**(1), 66–82 (2017)
20. Winker, S.K.: Quandles, knot invariants, and the n-fold branched cover. Ph.D. thesis, University of Illinois at Chicago (1984)

# A Typology of Mathematical Diagrams

Mikkel Willum Johansen[1](✉), Morten Misfeldt[2], and Josefine Lomholt Pallavicini[1]

[1] University of Copenhagen, Copenhagen, Denmark
mwj@ind.ku.dk, pallavicini@hum.ku.dk
[2] Aalborg University, Copenhagen, Denmark
misfeldt@learning.aau.dk

**Abstract.** In this paper, we develop and discuss a classification scheme that allows us to distinguish between the types of diagrams used in mathematical research based on the cognitive support offered by diagrams. By cognitive support, we refer to the gain that research mathematicians get from using diagrams. This support transcends the specific mathematical topic and diagram type involved and arises from the cognitive strategies mathematicians tend to use. The overall goal of this classification scheme is to facilitate a large-scale quantitative investigation of the norms and values governing the publication style of mathematical research, as well as trends in the kinds of cognitive support used in mathematics. This paper, however, focuses only on the development of the classification scheme.

The classification scheme takes its point of departure from case studies known from the literature, but in this paper, we validate the scheme using examples from a preliminary investigation of developments in the use of diagrams. Building on these results, we discuss the potential and pitfalls in using one generic classification scheme, as done in this analysis. This approach is contrasted with attempts that respect and build on individual diagram types, and as part of this discussion, we report the problems we experienced when using that strategy. The paper ends with a description of possible next steps in using text corpora as an empirical approach to understanding the nature of mathematical diagrams and their relation to mathematical culture.

**Keywords:** Classification of diagrams · Corpus analysis
Mathematical cognition

## 1 Introduction: Why Classify Mathematical Diagrams from a Cognitive Perspective

With this paper, we wish to develop a framework that allows us to characterize mathematical diagrams based on the kind of cognitive support they offer to mathematics. Furthermore, we aim to construct a classification scheme that bracket both the mathematical content represented by the diagram and the practice of specific mathematicians and to consider the diagram only as it presents itself on the page. Such a classification scheme will enable large-scale quantitative analysis of corpora of published mathematics papers that would otherwise be difficult to process.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 105–119, 2018.
https://doi.org/10.1007/978-3-319-91376-6_13

There are a number of good reasons for analyzing the use of diagrams in mathematical publications. The prominence of social norms governing the use of diagrams in publications was one result of an interview study with working mathematicians the two first authors of this paper conducted [1]. As a general trend, the mathematicians to whom we spoke considered various forms of diagrams and drawings to be vital to their work practices, but nevertheless, they tended to leave out such representations in their published papers.

Such results are hardly surprising to research mathematicians. Indeed, they generally confirm the point made nearly three decades ago by Hersh [2]: mathematics has a frontstage and a backstage; there is in mathematics a huge difference between the way a dish is prepared and the way it is presented to the public. However, recent work in the philosophy of mathematical practice indicates that the value and validity of inferences based on diagrams and figures should not be neglected, and the formalistic claim that such inferences are insecure and obsolete (e.g. [3: 43]) has been questioned by a number of scholars in the field [4, 5]. This vindication of diagrammatic reasoning was also confirmed in the interview study mentioned above as several of the interviewed mathematicians expressed the view that diagrams and other visual representations can constitute good arguments, and that it was a shame they were not allowed to be used in published papers.

Furthermore, investigations of historical cases have shown that diagrams and figures have played (and still play) a crucial role both in the conceptual development of mathematics [6–8] and in the framing and development of new areas of mathematics research [9, 10]. Such results might be seen as a call for a revision of the norms and values that govern publication practice. If inferences based in diagrams are (in some cases) valid, and if diagrams play a crucial part in the conceptual development of mathematics, perhaps we should allow diagrams in published papers to a greater extent.

To the working mathematician, the more pressing question is: are the norms and values of publication changing in mathematics? There are indications that we are in the midst of major changes in publication practice. In the year 1992 the mathematics journal *Experimental Mathematics* was launched with the direct intention to bring the behind-the-scene processes leading to mathematical discovery to the fore [11]. While *Experimental Mathematics* is mainly aimed at methodological developments connected to the use of computers as experimental tools, there are signs that a similar change in attitudes toward diagrams is underway. At least, diagrams seem to be brought more frequently to the frontstage of mathematics. It is not unusual for textbook expositions, even on advanced topics, to rely heavily on diagrams and drawings (e.g. [12]), and it also seems that mathematics journals contain more diagrams of still more varied types than they did at the height of the formalist movement in the middle of the 20th century. This, however, is merely an impression. We do not have substantial evidence to support the claim that the use of diagrams in mathematics journals is on the rise, and if it is, we do not know the shape or size of the change, when it started, or if all or only certain types of diagrams are being published more frequently. One way to address such questions in a thorough, empirical way is to track developments in the use of diagrams in mathematics research papers over time. This could be done by counting the number of diagrams included in research papers in selected influential mathematical journal over a long period of time (e.g., a century). However, in mathematics, a diagram is not simply a diagram. Diagrams are different and play varied cognitive roles in mathematical research; therefore, it would be interesting to track not only the number of diagrams but also the types of diagrams used.

To carry out such a large-scale empirical investigation of the use of diagrams in mathematics journals, an instrument that allows for counting and classifying diagrams with relative ease needs to be constructed. The objective of this paper is to describe and discuss a first attempt at building such an instrument. This leads to the following research question:

> *How can we classify mathematical diagrams published during the past century in a way that distinguishes among their different cognitive functions?*

To answer this question, we build on work by the first author of this paper. In [13] Johansen distinguishes between diagrams, symbols, and figures based in the different cognitive roles these representational forms play in mathematical practice. In the following, we introduce Johansen's distinction, adapt it by adding the category of Cartesian diagrams, and discuss the possible inclusion of a fourth category of matrices and tables. We demonstrate the classification scheme by describing how it worked and the problems we encountered when we used it in a systematic pilot investigation of the journal *Annals of Mathematics*. We should note that the focus of this paper is the development of the classification scheme, not developments in the use of diagrams. For this reason, we discuss only how well the classification scheme worked and not the preliminary results we obtained from using it.

## 2 Mathematical Diagrams and Figures

Although concepts such as diagram and figure are frequently discussed in the philosophy of mathematics, there is no consensus on precise definitions of the terms. Definitions range from the extreme, such as Peirce's [14: 90] functional definition of diagrams as any representation that allows the deduction of new information not used in its construction, to more restrictive definitions that come closer to the everyday use of the word, such as Larkin and Simon's [15] definition requiring a representation to involve an element of two-dimensionality to be considered a diagram.

Although Peirce's functional definition captures and justifies a central aspect of mathematical practice, namely, that mathematical work frequently involves translation between different representational types, it is too broad and does not allow for making cognitively and practically meaningful distinctions in the category of diagrams [16]. Moreover, Peirce's definition does not make it possible to distinguish between algebraic and diagrammatic representations (in the pre-theoretical sense of the words), so it is impossible to track the very changes in norms and publication practices that we are interested in here. Here, we are interested in actual mathematical practice and the cognitive role diagrams may play herein. For that reason, we use the less inclusive definition, which understands diagrams as representations that require two-dimensionality and uses this two-dimensionality in a non-trivial way (e.g. [15: 68, 17: 162]; see also [18] for a discussion). This definition comes closer to the everyday use of the word but also inherits some of the fuzziness of the pre-theoretical concept. When confronted with real-life examples of two-dimensional representations, choices have to be made to establish the precise boundaries of the concept, as we will see in the following.

Furthermore, when diagrams are defined in this way, the concept covers a rather inhomogeneous category of representations. To make useful distinctions in the concept of diagram, we will take departure in Johansen [13], where an attempt is made (a) to divide the class of diagrams into two subclasses (called 'diagrams' and 'figures') on the basis of cognitive function and (b) to explain how the cognitive functions provided by these two subclasses differ from those provided by symbols. It thus is pointed out that mathematical symbols can produce new knowledge by being subjected to purely syntactic manipulations. Some diagrams and figures may in the same way be subjected to purely syntactic manipulation, but what sets these representations apart from symbols from a cognitive perspective is that they can also be used for contentual and intuitive reasoning by relating mathematical objects to the everyday experience of the mathematician. The point in [13] is that members of the two subclasses of diagrams create this relation in different ways. One subclass (which Johansen calls 'figures') consists of representations such as Euclidian diagrams. When operating on a Euclidian diagram, the representation has a direct resemblance to the physical objects that constitute the abstraction class for the corresponding mathematical concept. When operating on a drawing of a triangle, for example, the representation has a direct resemblance to other perceptible triangles that make up the abstraction class for the general mathematical concept of 'triangle'. With the other subclass (which Johansen calls 'diagrams'), it is not possible to establish this kind of direct resemblance. Instead, diagrams display the mathematical objects they represent only if the objects are understood within a particular conceptualization. To give an example, a commutative diagram cannot be said to resemble neither the mathematical objects it represents (typically mathematical sets and maps between sets), nor the abstraction class of such objects. However, if the mathematical sets are conceptualized as objects located in space and the maps between them are conceptualized as movements between locations in space, the commutative diagram can be said to resemble such a situation. Similarly, Venn diagrams do not function via simple resemblance. The objects represented by such diagrams are typically not located in bounded regions of space and may not even have inherent spatial properties. The resemblance is only established it the objects are conceptualized as if they were bounded regions of space [c.f. 13: 100]. In other words, Johansen claims that diagrams in this second subclass function as material anchors for conceptual maps that relate mathematical content to concrete experiences in the physical world. We further refer to [19] for direct empirical evidence that research mathematicians in actual practice conceptualize the mathematical objects represented by diagrams in the way described (see also [17] for a further discussion of commutative diagrams' function as maps).

## 2.1 A Classification Scheme

In the following, we build on the basic idea from [13] but refine it. In particular, we do not follow Johansen in renaming a subclass of diagrams 'figures'. Instead, we introduce the following subclasses of diagrams:

1. Resemblance diagrams: This class roughly corresponds to Johansen's first sub-class of diagrams. That is, it includes diagrams that have a direct likeness to the physical objects that the corresponding mathematical concepts are supposed to model. This likeness can be geometric (as in Euclidian diagrams) or topological (as in knot diagrams).
2. Abstract diagrams: This class corresponds to Johansen's second class of diagrams. That is, it includes diagrams that are only meaningful if the mathematical content they represent is understood through a particular conceptual map.

In addition to these two subclasses, we will introduce a third subclass:

3. Cartesian diagrams: This subclass includes shapes and figures drawn in a coordinate system.

The reason for introducing this third subclass is largely pragmatic and grounded in our exploratory work with corpora of mathematical texts. The use of coordinate systems builds on conceptual blending allowing mathematical objects to simultaneously possess geometric and numerical properties [20: 385]. This function is not clearly included in the two categories introduced so far; therefore, we believe it necessary to propose this third subclass.

Furthermore, various forms of schemas, tables, and matrices can be said to constitute a fourth, independent category of two-dimensional visual inscriptions, and we discuss whether to include such representations as a fourth subclass. In this subclass, mathematical symbols are arranged in a way so that patterns in the physical layout of the representations may reveal mathematically relevant information (c.f. [13, 19]). This way of creating and using representations opens cognitive possibilities other than the three subclasses of diagrams described above, and such schemas fall outside the categories we have introduced. Tables and matrices, however, are not encompassed by the everyday concept of diagrams, and including such representations in the concept of diagrams might force us to broaden it further to include, say, the two-dimensional arrangement of numeral symbols in pen-and-paper multiplication. Here, therefore, a choice will clearly have to be made.

## 2.2 Applying the Scheme

The motivation for the development of the classification scheme is the need to carry out large-scale, empirical investigations of the use of diagrams in mathematics journals. Applying the scheme for such a purpose leads to a trade-off between reliability and quantity: on one hand, reliable classification of a given diagram in the scheme requires in-depth case studies exploring how mathematicians conceive of and use the diagram in actual practice, but on the other hand, an investigation of the general trends in diagram use requires the classification of large quantities of diagrams. Because the suggested classification scheme takes the way mathematicians perceive diagrams as its point of departure we believe that an examination of the typographical features of a given diagram and a minimum of mathematical context will give a clear indication of how the diagram should be classified in the scheme. Of course, misclassifications are inevitable, but we are primarily interested in general trends in the frequency and types of diagrams used, so a few misclassifications will not be detrimental to achieving the overall objective of the investigation. However, it is clear that transparency and reflections about reliability and error margins must be a constant, important concern in quantitative analysis of this type.

Finally, it should be noted that the move to quantitative analysis of diagram use suggested here is new in the philosophy of mathematical diagrams. In certain respects, it stands in contrast to core values of the paradigm, where careful case studies are the norm. We do hope the reader will appreciate the possible benefits of deviating from this norm.

## 3 Examples

In the following, we give a few representative examples of the diagrams we encountered during our pilot study on the diagrams used in *Annals of Mathematic*. The selection represents the different kinds of diagrams we found by investigating the journal over ten-year intervals in the period from 1885 to 2005. For each year (e.g. 1885, 1895), we analyzed all the research papers published in the journal, marked all the two-dimensional representations, and tried to apply the classification scheme to them. All the diagrams we present in the following sections are reproduced with kind permission from *Annals of Mathematics*.

### 3.1 Prototypical Diagrams

Most diagrams we encountered were easily classified using the suggested scheme. We begin with three prototypical examples illustrating how the subclasses of the scheme function when confronted with representations from real-life publications.

As a first example, we consider a figure from [21] (Fig. 1). Notice that the representation is classified as a figure in the original text (and not as a diagram, Euclidian diagram, or similar terms). This illustrates the confusion in nomenclature discussed above. In the classification scheme suggested here, however, the representation clearly belongs to the subclass of resemblance diagrams as it displays a circle and geometric constructions involving chords within the circle. From a cognitive point of view, the diagram supports reasoning with these objects by anchoring the conceptual structure in a stable medium (the paper) so that the mathematical objects are represented by shapes that geometrically resemble the class of physical objects modelled by the mathematical concepts. In other words, in this case, our categories allow for a clear and obvious classification of the diagram.

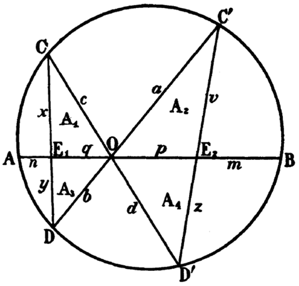


**Fig. 1.** A resemblance diagram. Reproduced with permission from [21: 175].

As our second example, we consider a representation from [22] (Fig. 2). In this case, the representation is not labeled, but the accompanying text refers to it as a commutative diagram.

$$\begin{tikzcd}
0 \arrow[r] & C(\mathfrak{V}, \mathfrak{A}) \arrow[r] \arrow[d, "\tau"'] & C(\mathfrak{V}, \mathfrak{B}) \arrow[r] \arrow[d, "\tau"'] & C(\mathfrak{V}, \mathfrak{C}) \arrow[d, "\tau"'] \\
0 \arrow[r] & C(\mathfrak{U}, \mathfrak{A}) \arrow[r] & C(\mathfrak{U}, \mathfrak{B}) \arrow[r] & C(\mathfrak{U}, \mathfrak{C})
\end{tikzcd}$$

**Fig. 2.** An abstract diagram. Reproduced with permission from [22: 216].

We consider the diagram to belong to the class of abstract diagrams and claim that, in contrast to the diagram in Fig. 1, this diagram does not directly resemble any aspect of physical reality modelled by the mathematical objects involved. Rather, the diagram only resembles the mathematical objects if they are viewed from a particular conceptualization. We, however, refrain from identifying or describing the conceptualization because that requires the kind of in-depth contextual analysis we are trying to avoid here (see [13, 19] for examples).

As a third example, we look at the following representation from [23] (Fig. 3):

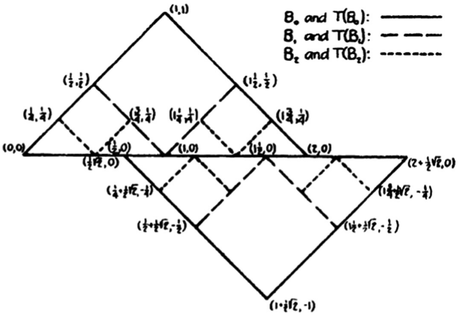


**Fig. 3.** A Cartesian diagram. Reproduced with permission from [23: 43].

This diagram involves a construction in a coordinate system, and we interpret it as representing a conceptual blend, where points are simultaneously considered to be locations on the geometric plane and elements in a set of ordered pairs of numbers ($x$, $y$). In other words, we see the diagram as belonging to the subclass of Cartesian diagrams.

## 3.2 Is This a Diagram?

Most of the diagrams we inspected fall neatly into one of the three categories of the classification scheme, but we also encountered diagrams challenging the scheme. In this and the following section, we discuss two types of challenges: cases where the representations in question are on the borderline of being considered a (mathematical) diagram at all and cases where a diagram is on the borderline of two of the subclasses in our classification scheme.

Beginning with the first type of concern, we return to the commutative diagram presented in Fig. 2. Interestingly, this diagram is preceded by the following representation (Fig. 4):

$$\cdots \to H^q(\mathfrak{U}, \mathfrak{B}) \to H^q_0(\mathfrak{U}, \mathfrak{C}) \xrightarrow{d} H^{q+1}(\mathfrak{U}, \mathfrak{A}) \to H^{q+1}(\mathfrak{U}, \mathfrak{B}) \to \cdots,$$

**Fig. 4.** Exact sequence. Reproduced with permission from [22: 216].

In many ways, this representation seems similar to the diagram presented in Fig. 2. Letters representing mathematical objects are connected with arrows representing maps between these objects. Why is this not a diagram? The answer is simply that the representation is not two dimensional but one dimensional (in the sense that it can be read linearly). Although this distinction between one- and two-dimensional representations may seem arbitrary, it, at least in this case, connects to a real distinction in the language used in mathematical practice. In the text, Fig. 2 is referred to as a (commutative) diagram, while the representation in Fig. 4 is referred to as something different: an exact sequence. This furthermore illustrates a fundamental difficulty connected to delaminating diagrams as a distinct type of representations. Functional categories, such as the one introduced by Peirce [14] or the one attempted in [13], tend to clash with the pre-theoretical language used by practitioners. From a cognitive perspective, it is very difficult to point to the operative difference between an exact sequence and a commutative diagram, but in mathematical practice, the two representations appear to have different statuses, at least judging from the language (one is labeled a diagram, and the other is not). It is our goal in this paper to create an instrument that makes it possible to track and understand aspects of mathematical practice, so we have to make pragmatic compromises. One such compromise is to adopt the criterion that a diagram has to be two dimensional, even if this criterion may seem arbitrary from a theoretical perspective (as the classification of commutative diagrams as diagrams and exact sequences as non-diagrams illustrates).

As a second example of a representation that is on the borderline of being considered a mathematical diagram, we can look at the following representation (Fig. 5):

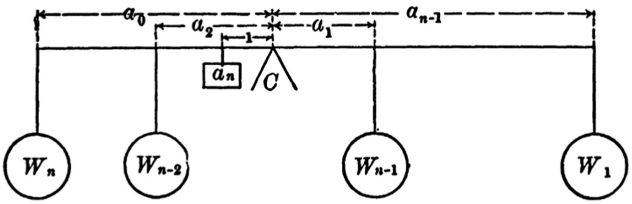


**Fig. 5.** Cartesian diagram. Reproduced with permission from [24: 73].

The circles marked $W_n$ to $W_1$ represent water-filled vessels, $C$ represents a fulcrum, and the overall diagram represents a mathematical problem expressed in terms of balancing weights around a fulcrum. The diagram pictures the structure of an (imagined) physical mechanism. As a first approximation, it might be taken to be a resemblance diagram, similar to Euclidian diagrams picturing geometric structures. However, mathematical resemblance diagrams typically involve only geometric and/or topological features of the represented objects, whereas the diagram in Fig. 5 also involves physical features, such as weight, equilibrium, and movement. For this reason, it is questionable whether the diagram is a *mathematical* diagram. Perhaps it is a physics diagram, which might not be the same as those used in mathematics. In other words, the diagram in Fig. 5 challenges us to draw another (more or less arbitrary) border between the diagrams used in mathematics and those used in physics. Although the diagram is on the borderline, we need a clear border; we would not consider, say, Feynman diagrams to be mathematical diagrams, although they are clearly diagrams.

During the pilot study, we also frequently encountered schemas and matrices such as that reproduced in Fig. 6:

$$(C^{\Delta})^{-1} = \frac{1}{x\delta}\begin{pmatrix} 1 & -x & 0 & 0 & \cdots \\ -x & 1+x^2 & -x & 0 & \cdots \\ 0 & -x & 1+x^2 & -x & \cdots \\ & & & \ddots & -x \\ & \cdots & 0 & -x & 1 \end{pmatrix}.$$

**Fig. 6.** Matrix. Reproduced with permission from [25: 198].

As noted, it can be argued from a cognitive perspective that such representations play a different role in mathematical practice than diagrams, in particular, they allow abstract mathematical structures to be visualized as patterns in the physical arrangements of symbols on the paper. However, although such representations clearly exploit their two-dimensionality, they are traditionally not considered to be diagrams, and for that reason, we decided not to include them in our pilot study. This can again be seen as a somewhat arbitrary choice. However, as argued, including matrices and similar schemas in our category of diagrams would introduce a demarcation problem: if we consider matrices and schemas to be diagrams, we would also have to consider, say, the arrangement of numerals used for pen-and-paper multiplication as diagrams. Although this might make sense from a purely functional perspective, important aspects of the distinction between mathematical discourse and diagrams would cease to exist. The distinctions we consider here are based on both the two-dimensional aspect of diagrams and established practice among mathematicians, which usually does not consider matrices and schemas to be diagrams. The example of matrices and schemas, however, clearly and interestingly illustrates once again that the pre-theoretical concept of diagrams found in mathematical practice does not correspond with purely functional or theoretical categories.

### 3.3 What Kind of Diagram Is This? Examples that Challenge the Classification Scheme

Another type of challenge came from diagrams that did not fall clearly into the categories in the classification scheme. We begin by considering the following representation from [24] (Fig. 7):

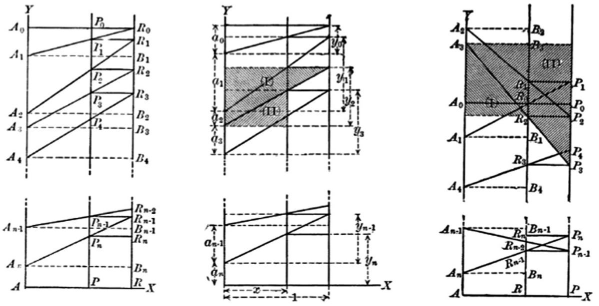


**Fig. 7.** Cartesian diagram. Reproduced with permission from [24: 66].

We see this as a Cartesian diagram, but the identification of the diagram as something involving a coordinate system is less straightforward than in the prototypical example above (Fig. 3). The construction of unity in the part of the diagram marked “Fig. 2” indicates that we are dealing with a Cartesian construction, but to make sure that it is not a resemblance diagram, we have to consider the textual context. Here, it is made clear that the diagram indicates the pointwise construction of the graph of a polynomial. It, therefore, is most sensible to classify the diagram as Cartesian because the overall framework of the diagram involves a Cartesian blend.

As a second example of a diagram challenging the subclasses of our scheme, we can look at the following representation (Fig. 8) from [26]:

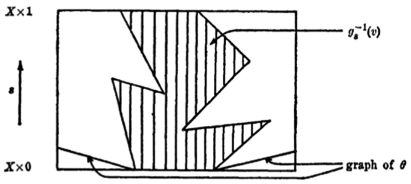


**Fig. 8.** Abstract diagram. Reproduced with permission from [26: 107].

Once more, the categorization is not straightforward, but as an educated guess, we consider the diagram to be what we call an abstract diagram. The reason for this categorization is partly positive and partly negative. The diagram does seem to involve abstract elements, such as sets represented as locations in space, and although it also includes geometric elements, we do not believe these to resemble the mathematical objects in any straightforward way. In other words, the diagram seems to be conceptualized, although a contextual analysis is necessary to identify the exact conceptualization. A similar analysis can be extended to other diagrams, such as the following (Fig. 9):

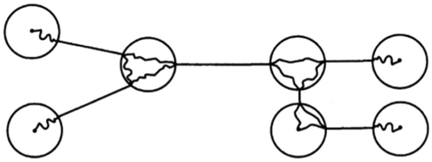


**Fig. 9.** Abstract diagram. Reproduced with permission from [27: 223].

As our final example, we look at two diagrams from [28]. The first one is the following diagram (Fig. 10):

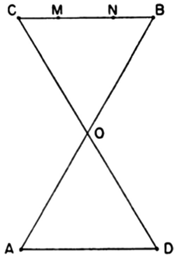


**Fig. 10.** Resemblance diagram. Reproduced with permission from [28: 609].

As we see it, this is a resemblance diagram displaying a particular situation. If we look at the textual context, we furthermore can see that, in this case, the resemblance is topological, not geometric, although the particular type of resemblance is not consequential in our classification scheme. However, if we continue to read the paper [28], we encounter the following figure (Fig. 11):

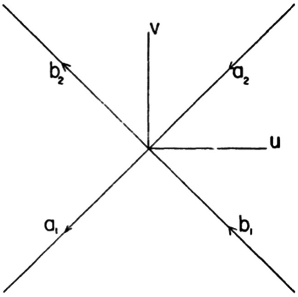


**Fig. 11.** Resemblance diagram. Reproduced with permission from [28: 613].

If we know that we are dealing with topology, it is not difficult to decode the diagrams as resemblance diagrams (based on topological resemblance), but without such knowledge, we believe it would be very difficult to make a meaningful classification of these diagrams in our scheme. In difficult cases such as this, it is, in other words, not possible to completely bracket the intellectual context of the diagram and the classification cannot be more than an educated guess.

## 4 Discussion: Balancing the Resolution of Diagram Classification

Our experience from the pilot study indicates that the classification scheme we have presented makes it possible to handle most of the diagrams published over a century in a leading mathematical journal. Although most classifications are straightforward, we also encountered cases where we had to involve the textual or intellectual context of diagrams to classify them, and in other cases, we could only give educated guesses.

It, therefore, might be worthwhile contemplating possible alternatives. When we began the pilot study, we initially set out to use a classification based on the standard names given to various diagram types (e.g., commutative diagrams, Dynkin diagrams, and simplex diagrams). Although such a classification gave a more fine-grained resolution in the diagram classification—which allowed tracking trends in the use of specific diagram types—it turned out to be difficult to carry out in practice, at least on a large scale. Mathematicians do not consistently name diagrams, and often diagrams are not referred to in the text by their standard names. Consequently, counting the number of diagrams of a specific type requires making judgements based on the visual appearance of diagrams to classify them correctly. Furthermore, although mathematicians tend to express themselves using a relatively small number of standard types of diagrams, they also occasionally use their own idiosyncratic representations. Although such cases are difficult to handle in the proposed classification scheme, they are in principle out of the reach of a classification taking departure in standard names.

In another alternative, one might consider counting the number of diagrams without attempting to classify them further. This would reduce some of the uncertainty we encountered, but at the price of a similar reduction in information. The subclasses in our classification scheme delineate three very different kinds of diagrams, ranging from direct geometric representations (e.g., Euclidian diagrams) to completely abstract representations (e.g., commutative diagrams). The ability to track the development not only of the frequency of diagrams but also of the relative frequency of the different subclasses of diagrams could yield valuable information about the norms governing the use of diagrams in mathematical practice. Furthermore, even a simple count of the number of diagrams without attempting to classify them further, is not without uncertainty. As we have seen, the concept of mathematical diagrams is not well defined, and many of the challenges we faced concerned not how to classify diagrams within the scheme but whether to count a given representation as a mathematical diagram.

Finally, it should be noted that the classification of diagrams can be performed using dimensions other than those suggested here, such as Stenning's [18] distinction between directness and indirectness or de Toffoli's [17] distinctions among expressiveness, calculability, and transparency. Our choice of classification scheme is mainly pragmatic. We believe that it will work because its point of departure is mathematicians' conception of diagrams and for that reason it will provide relevant information about the changes and trends in the use of diagrams in mathematics publications. This, of course, does not rule out the possibility of including other dimensions in the categorization and quantitative analysis of diagram use. Furthermore, we cannot rule out that the classification scheme will have to be augmented or adjusted in order to be applied to other corpora of mathematical texts; we have tested the scheme only on a single journal.

## 5 Conclusion

In this paper, we have argued for a change in the philosophy of mathematical diagrams. We believe it will be productive to augment in-depth, qualitative case-studies of diagrams with quantitative investigations tracking overall trends in diagram use. We have developed the first version of an instrument that can be used to classify diagrams in large-scale, quantitative studies, and we have demonstrated the function of the instrument by reporting on a pilot investigation on a mathematical journal. The instrument generally seems to give reliable results, but as we have also seen, the quantitative approach proposed here also leads to a number of dilemmas. A choice between the reliability of the investigation and the relevance of the generated knowledge must be made. Furthermore, the pre-theoretical concept of mathematical diagram is not well defined, and choices have to be made between the use of a well-defined theoretical concept of diagrams and the ability to track and speak into actual mathematical practice and mathematicians' understanding of this practice. However, as we see it—and as our pilot study confirms—the challenges confronting a quantitative approach are manageable, and the full-scale application of the proposed instrument has the potential to generate valuable information about long-term developments in the norms governing the use and publication of mathematical diagrams.

## References

1. Johansen, M.W., Misfeldt, M.: An empirical approach to the mathematical values of problem choice and argumentation. In: Larvor, B. (ed.) Mathematical Cultures. THS, pp. 259–269. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-28582-5_15
2. Hersh, R.: Mathematics has a front and a back. Synthese **80**(2), 127–133 (1991)
3. Pasch, M., Dehn, M.: Vorlesungen über Neuere Geometrie. Die Grundlehren der Mathematischen Wissenschaften, vol. 23. Springer, Berlin (1926)
4. Brown, J.R.: Philosophy of Mathematics, an Introduction to a World of Proofs and Pictures. Philosophical Issues in Science. Routledge, London (1999)
5. Giaquinto, M.: Visual Thinking in Mathematics, An Epistemological Study. Oxford University Press, New York (2007)
6. Livingston, E.: Cultures of proving. Soc. Stud. Sci. **29**(6), 867–888 (1999)
7. Krömer, R.: Tool and Object, a History and Philosophy of Category Theory. Science Networks Historical Studies, vol. 32. Birkhäuser, Basel (2007)
8. Carter, J.: Diagrams and proofs in analysis. Int. Stud. Philos. Sci. **24**(1), 1–14 (2010)
9. Kjeldsen, T.H.: From measuring tool to geometrical object: Minkowski's development of the concept of convex bodies. Arch. Hist. Exact Sci. **62**(1), 59–89 (2008)
10. Johansen, M.W., Misfeldt, M.: Semiotic scaffolding in mathematics. Biosemiotics **8**, 325–340 (2015)
11. Epstein, D., Levy, S., Llave, R.: About this journal. Exp. Math. **1**(1), 1–3 (1992)
12. Hatcher, A.E.: Topology of Numbers. https://www.math.cornell.edu/~hatcher/TN/TNbook.pdf. Accessed 4 Dec 2017
13. Johansen, M.W.: What's in a diagram? On the classification of symbols, figures and diagrams. In: Magnani, L. (ed.) Model-Based Reasoning in Science and Technology. Theoretical and Cognitive Issues (SAPERE), vol. 8, pp. 89–108. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37428-9_6
14. Stjernfelt, F.: Diagrammatology, an Investigation on the Borderlines of Phenomenology, Ontology, and Semiotics. Synthese Library, vol. 336. Springer, Dordrecht (2007). https://doi.org/10.1007/978-1-4020-5652-9
15. Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. Cogn. Sci. **11**(1), 65–100 (1987)
16. Steensen, A.K., Johansen, M.W.: The role of diagram materiality in mathematics. Cogn. Semiot. **9**(2), 183–201 (2016)
17. De Toffoli, S.: 'Chasing' the diagram—the use of visualizations in algebraic reasoning. Rev. Symb. Log. **10**(1), 158–186 (2017)
18. Stenning, K.: Distinctions with differences: comparing criteria for distinguishing diagrammatic from sentential systems. In: Anderson, M., Cheng, P., Haarslev, V. (eds.) Diagrams 2000. LNCS (LNAI), vol. 1889, pp. 132–148. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-44590-0_15
19. Johansen, M.W., Misfeldt, M.: Material representations in mathematical research practice (in preparation)
20. Lakoff, G., Núñez, R.: Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being. Basic Books, New York (2000)

21. Candy, A.L.: A general theorem relating to transversals, and its consequences. Ann. Math. **10**(1), 175–190 (1895)
22. Serre, J.-P.: Faisceaux algebriques coherents. Ann. Math. **61**(2), 197–278 (1955)
23. Gehman, H.M.: Concerning the subsets of a plane continuous curve. Ann. Math. **27**(1), 29–46 (1925)
24. Moritz, R.E.: Some physical solutions of the general equation of the $n^{th}$ degree. Ann. Math. **6** (3), 64–78 (1905)
25. Guerra, F., Rosen, L., Simon, B.: The $P(\varphi)_2$ Euclidean quantum field theory as classical statistical mechanics. Ann. Math. **101**(2), 191–259 (1975)
26. Hatcher, A.E.: Higher simple homotopy theory. Ann. Math. **102**(1), 101–137 (1975)
27. Sela, Z.: The isomorphism problem for hyperbolic groups I. Ann. Math. **141**(2), 217–283 (1995)
28. Morse, M., Heins, M.: Topological methods in the theory of functions of a single complex variable I. Deformation types of locally simple plane curves. Ann. Math. **46**(4), 600–624 (1945)

# The Classificatory Function of Diagrams: Two Examples from Mathematics

Christophe Eckes[1] and Valeria Giardino[2](✉)

[1] Archives Henri-Poincaré - Philosophie et Recherches sur les Sciences et les
Technologies/Université de Lorraine, Nancy, France
christophe.eckes@univ-lorraine.fr

[2] Archives Henri-Poincaré - Philosophie et Recherches sur les Sciences et les
Technologies/CNRS, Nancy, France
valeria.giardino@univ-lorraine.fr

**Abstract.** In a recent paper, De Toffoli and Giardino analyzed the practice of knot theory, by focusing in particular on the use of diagrams to represent and study knots [1]. To this aim, they distinguished between *illustrations* and *diagrams*. An illustration is *static*; by contrast, a diagram is *dynamic*, that is, it is closely related to some specific inferential procedures. In the case of knot diagrams, a diagram is also a well-defined mathematical object in itself, which can be used to classify knots. The objective of the present paper is to reply to the following questions: Can the classificatory function characterizing knot diagrams be generalized to other fields of mathematics? Our hypothesis is that dynamic diagrams that are mathematical objects in themselves are used to solve classification problems. To argue in favor of our hypothesis, we will present and compare two examples of classifications involving them: (i) the classification of compact connected surfaces (orientable or not, with or without boundary) in combinatorial topology; (ii) the classification of complex semisimple Lie algebras.

**Keywords:** Diagrammatic reasoning · Knot diagrams
Mathematical classifications · Combinatorial topology
Compact connected surfaces · Lie algebras

## 1 Introduction: Diagrams and Classifications

In a recent paper, De Toffoli and Giardino discussed the practice of a particular branch of topology, knot theory, by focusing in particular on the use of diagrams to represent and study knots [1]. A knot $K$ in tridimensional space is a submanifold that is diffeomorphic to a circle. The main aim of knot theory consists in classifying the different knots that are embedded in ambient tridimensional space. Two knots $K$ and $K'$ in $\mathbb{R}^3$ are *ambient isotopic* if there exists a *continuous map* $h : [0, 1] \times \mathbb{R} \to \mathbb{R}^3$ such that $h_0 = \mathrm{Id}_{\mathbb{R}^3}$, $h_1(K) = K'$ and $h_t$ is a diffeomorphism. A knot $K$ (embedded in $\mathbb{R}^3$) can be represented by

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 120–136, 2018.
https://doi.org/10.1007/978-3-319-91376-6_14

means of a *knot diagram*, which is a particular projection of $K$ onto $\mathbb{R}^2$. The authors proposed to distinguish between *illustrations* and *diagrams*. An illustration is a *static* representation that can of course convey some information about what it represents, in this case a knot type; however, its possible transformations are not well-defined. By contrast, a diagram is a *dynamic* representation, that means, a representation that is closely related to some specific inferential procedures. Differently from mere illustrations, diagrams would not only suggest effective problem-solving strategies, but, more interestingly, they would end up being themselves part of the proof of a mathematical result. A knot diagram is a dynamic *tool* because an expert imagines performing some movements on it according to some specific inferential procedures. However, it is also a *mathematical object* in itself, which is well-defined and intended to solve a classification problem. Knot diagrams have thus a *classificatory function*: they determine the classification of knots via ambient isotopies and at the same time they are themselves the objects of such a classification.

The objective of the present paper is to reply to the following questions: Can the classificatory function characterizing knot diagrams be generalized to other parts of mathematics? To clarify, our hypothesis is that if we intend diagrams as dynamic tools allowing for some inferential procedures that are at the same time mathematical objects, then we can find other cases of diagrams having a classificatory function. To argue in favor of our hypothesis, in the remainder of the paper we will present two examples of classifications that involve the use of diagrams. In Sect. 2 we will consider combinatorial topology and discuss the use of diagrammatic and symbolic representations for the classification of compact connected surfaces (orientable or not, with or without boundary), ranging from Dyck's diagrams to Alexander and Hilbert fundamental polygons. In Sect. 3, we will turn to algebraic structures and consider the diagrammatic representations that were introduced for the classification of complex semisimple Lie algebras, ranging from Artin-van der Waerden's vector diagrams to Dynkin diagrams. In Sect. 4, we will discuss and compare these two examples, and draw some conclusions.

## 2 The Classification of Compact Surfaces

In this section, we focus on the classification of topological surfaces[1] (with or without boundary) that are compact (each of their open covers has a finite subcover). First, we will very briefly recall the main historical steps that led to this classification. In particular, we will mention the special cases of compact *Riemann* surfaces and of compact orientable *topological* surfaces (without boundary). Second, we will introduce Dyck's diagrams, which allow for a complete treatment of the general case (i.e. topological surfaces with or without boundary, orientable or not). Finally, we will describe another diagrammatic representation related to the classification of such objects.

---

[1] A surface is supposed to be a connected two-dimensional topological manifold, that is it cannot be represented as the union of two disjoint nonempty open sets.

### 2.1 First Classification

The interest for the properties of *analysis situs* that some surfaces imply arose with the publication of Bernhard Riemann's dissertation (1851) and of his memoir on Abelian integrals (1857). Riemann was interested in making algebraic functions of a complex variable uniform, and to this aim he introduced a particular class of Riemann surfaces[2] (compact Riemann surfaces), which in the end became his main research object. In particular, to classify them, he conceived some "paradigmatic" surfaces (the sphere and the tori with $g$ holes) and individuated a fundamental topological invariant: the genus $g$. Two Riemann surfaces of the same genus that are holomorphically equivalent are also topologically equivalent; however, the classification implies that the opposite is not always true. A crucial result states that every compact surface without boundary and orientable is homeomorphic to the sphere $\mathbb{S}^2$ (which is simply connected) or to the connected sum of $g \geqslant 1$ tori $\mathbb{T}^2$ (with $g$ being the genus of such a surface). The classification so obtained amounts to reducing all compact surfaces without boundary to paradigmatic surfaces that are pairwise non-homeomorphic.

However, some compact surfaces without boundary are non-orientable. There are two fundamental examples:

1. The *real projective plane* $\mathbb{P}_2(\mathbb{R}) = (\mathbb{R}^3 \backslash \{0\})/\mathscr{R}$ where $\mathscr{R}$ designates the relation "to be collinear with", which is homeomorphic to $\mathbb{S}^2 \backslash (x \sim -x)$, i.e. to the quotient of the sphere of $\mathbb{R}^3$ by antipodal relation.[3]
2. The *Klein bottle* $K$, described by Klein in 1882.[4]

The representations of such surfaces in $\mathbb{R}^3$ are characterized by auto-intersections, which means that they cannot be embedded in $\mathbb{R}^3$.

### 2.2 Dyck's Classification

A first classification of compact surfaces with or without boundary (no matter if orientable or not) is due to a disciple of Klein, Walther von Dyck, who put forward two kinds of related representations for non-orientable surfaces with or without boundary [4]: the "fundamental forms" (*Grundformen*) in $\mathbb{R}^2$, which display the instructions on how to construct the "normal forms" (*Normalformen*) in $\mathbb{R}^3$. These representations allow for the classification of non-orientable surfaces whose genus $g$ is between 1 and 3.

Following Dyck's numbering in the table shown in Fig. 1, Fig. 5 corresponds to the real projective plane, with the Steiner surface as normal form, which is a mapping of $\mathbb{P}_2(\mathbb{R})$ into $\mathbb{R}^3$; Fig. 6 corresponds to the Möbius band and Fig. 7 to the Klein bottle. Therefore, Figs. 5 and 7 represent two non-orientable

---

[2] A Riemann surface is an *analytic* manifold of (complex) dimension 1. From the point of view of the *analysis situs*, a Riemann surface is orientable.

[3] See for reference [2, pp. 550–551].

[4] [3, p. 80]. From this booklet, it does not seem that Klein has used any figure to illustrate this description.

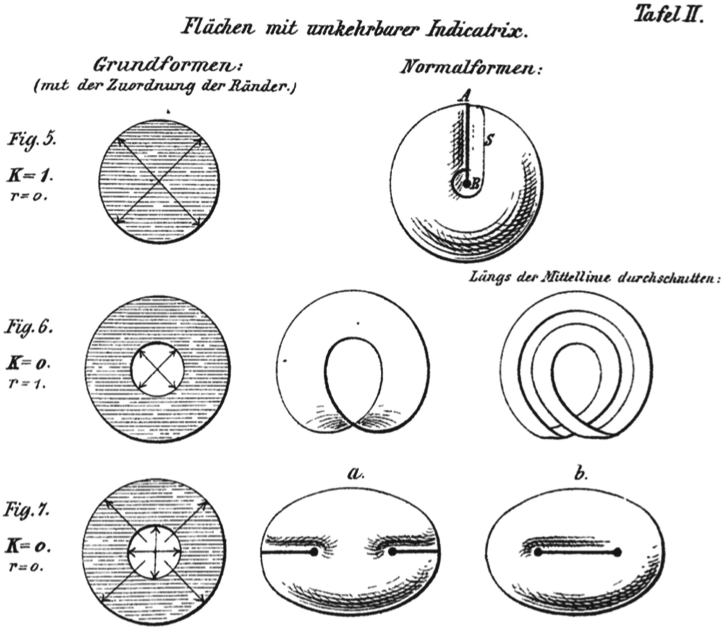


**Fig. 1.** Dyck's diagrams from Fig. 1, to Fig. 5 in Tafel II in *Mathematische Annalen*, volume 32.

surfaces without boundary, i.e. where the number of boundary components is $r = 0$. The number $K$ designates in Dyck's table the Euler characteristic $\chi(S)$. In the case of a compact, non-orientable surface, $\chi(S)$ is given by the formula $\chi(S) = 2 - g$, where $g$ is the genus of $S$: (i) the real projective plane $\mathbb{P}_2(\mathbb{R})$ is of genus 1, and therefore $\chi(\mathbb{P}_2(\mathbb{R})) = 1$; (ii) the Klein bottle is of genus 2, and therefore in this case $\chi(S) = 0$. If $S$ admits a boundary, then $\chi(S) = 2 - r - g$, where $r$ is the number of boundary components. The Euler characteristic of the Möbius band is therefore equal to zero, because in this case $g = 1$ and $r = 1$. The two normal forms $a$ and $b$ associated to Fig. 7 allow considering the Klein bottle as the "connected sum" of two copies of the real projective plane. Following now Dyck's numbering in the table shown in Fig. 2, Figs. 8 and 9 correspond to surfaces with boundary, and Fig. 10 to Dyck's surface, which has no boundary. The (non-orientable with boundary) surface corresponding to Fig. 8 admits $r = 2$ boundary components. It is of genus 1 and in this case $\chi(S) = -1$. The (non-orientable with boundary) surface corresponding to Fig. 9 admits $r = 1$ boundary components. It is of genus 2 and $\chi(S) = -1$. The (non-orientable without boundary) surface corresponding to Fig. 10 is of genus $g = 3$ and $\chi(S) = -1$. The three normal forms $a$, $b$, $c$ show that this surface $D$ ("D" for "Dyck's surface") corresponds to: (i) the connected sum of a torus and $\mathbb{P}_2(\mathbb{R})$; (ii) the connected sum of a Klein bottle and $\mathbb{P}_2(\mathbb{R})$; (iii) the connected sum of three copies of $\mathbb{P}_2(\mathbb{R})$.

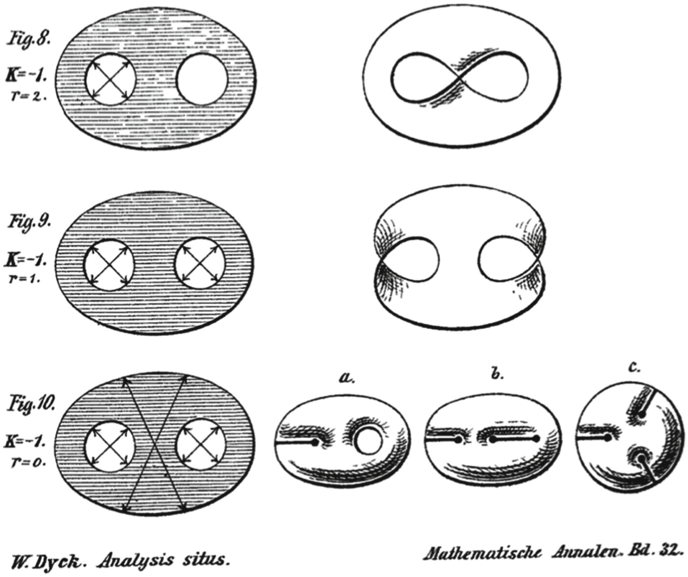


**Fig. 2.** Dyck's diagrams from Fig. 6 to Fig. 10 in Tafel II in *Mathematische Annalen*, volume 32.

These representations are all well-defined. In the case of the *Grundformen*, the arrows indicate that one has to identify the points that are diametrically opposed. For example, the real projective plane can be interpreted as a hemisphere such that the points that are diametrically opposed on its equator are identified. Moreover, a continuous deformation of this hemisphere leads us back to the closed disk $\overline{D} = \{(x, y) \in \mathbb{R} \mid x^2 + y^2 \leqslant 1\}$. The real projective plane is thus a closed disk such that the points that are diametrically opposed on its boundary are identified: this procedure corresponds exactly to the *Grundform* as presented in Fig. 5. In the case of compact non-orientable surfaces that have no boundary, a fundamental "building block" clearly emerges: the real projective plane $\mathbb{P}_2(\mathbb{R})$. In fact, $K \cong \mathbb{P}_2(\mathbb{R})\#\mathbb{P}_2(\mathbb{R})$ and $D \cong \mathbb{T}^2\#\mathbb{P}_2(\mathbb{R}) \cong K\#\mathbb{P}_2(\mathbb{R}) \cong \mathbb{P}_2(\mathbb{R})\#\mathbb{P}_2(\mathbb{R})\#\mathbb{P}_2(\mathbb{R})$,[5] where # is the connected sum.[6] Dyck gets to the following result [4, p. 488]: two compact surfaces are *homeomorphic* if and only if: (a) they have the same Euler characteristic; (b) they are both orientable or both non-orientable; (c) they have the same boundary component $r$ (that can be put aside when the surfaces in question have no boundary). As a consequence, non-orientable surfaces corresponding to Figs. 5, 6, 7, 8, 9 and 10 are pairwise

---

[5] Note that the equivalence $\mathbb{T}^2\#\mathbb{P}_2(\mathbb{R}) \cong K\#\mathbb{P}_2(\mathbb{R})$ does not entail $\mathbb{T}^2 \cong K$. In fact, $\mathbb{T}^2$ is orientable while $K$ is not.

[6] The *connected sum* of two surfaces $\Sigma_1$ and $\Sigma_2$, is obtained by deleting from $\Sigma_1$ a disk with boundary circle $c_1$ and from $\Sigma_2$ a disk with boundary circle $c_2$ before regluing $c_1$ and $c_2$. In particular, $\chi(\Sigma_1\#\Sigma_2) = \chi(\Sigma_1) + \chi(\Sigma_2) - 2$, a formula which allows inferring that $\chi(K) = 0$, by knowing that $\chi(\mathbb{P}_2(\mathbb{R})) = 1$.

non-homeomorphic. Dyck shows with no difficulties that two surfaces that do not satisfy at the same time conditions $(a)$, $(b)$ and $(c)$ are non-homeomorphic.[7]

The representations used by Dyck satisfy the properties of a mathematical diagram as we intend it: they are well-defined, dynamic, and associated to some inferential procedures. However, they must be interpreted in relation to a *text*, without which it would be impossible to understand the meaning of the constants $K$ and $r$ (in Dyck's notation) that are crucial for the organization of the classification, going from $K = 1$ to $K = -1$[8]. Moreover, they are arranged in a table that encourages us to compare them: if one reads Dyck's table *from left to right*, the relation between *Grundform* and *Normalform*, but also between equivalent *Normalformen* is seen[9]; if one reads Dyck's table *from above to below*, (i) the real projective plane appears as the fundamental building block for Figs. 7 and 10, thus leading us from simpler to more complex compact, non-orientable surfaces without boundary and (ii) the surfaces are recognized as pairwise non-homeomorphic, i.e. there is no redundancy. The diagrammatic representations used by Dyck reveal that his aim was to find a solution to the classification problem for surfaces in general, i.e. for compact surfaces with or without boundary, orientable or not. The case of the real projective plane, initially problematic because of the absence of a representation without auto-intersection in $\mathbb{R}^3$, becomes now *generic*: it is the building block thanks to which all compact non-orientable surfaces without boundary can be constructed and classified.

### 2.3 The "Fundamental Polygon"

In this section we will present another diagrammatic representation for compact surfaces that was introduced independently from Dyck's diagrams and made widely popular by Seifert and Threlfall [5]. In their monograph on topology, the authors showed that a compact surface can be reduced to its Poincaré fundamental polygon (*Poincarésche Fundamentalpolygon*). To clarify, this expression is to some extent misleading. In fact, Poincaré [6, p. 47] was not primarily concerned with the classification of compact surfaces; his main goal was rather to show how a three-dimensional topological manifold could be represented by a family of polyhedra. However, in order to illustrate such a representation to the reader, he referred to surfaces and described the particular procedure of obtaining a torus from a rectangle, by identifying its opposite sides in a specific way. In his words:

> Consider a rectangle $ABCD$ and a torus on which we draw two cuts, namely, latitudinal and longitudinal circles; let $H$ be their point of intersection. The surface of the torus will then be homeomorphic to the rectangle; the two sides of the cut formed by the longitudinal circle correspond

---

[7] Dyck conjectured but did not prove that all compact surfaces can be reduced to the normal forms he had identified.

[8] For instance, in the case of Figs. 6 and 7, $K = 0$ but $r = 1$ for Fig. 6 and $r = 0$ for Fig. 7; therefore, the two surfaces are non-homeomorphic.

[9] Note that the figure next to the *Normalform* of the Möbius band is a Möbius band cut along its middle line. We are not sure why Dyck put this figure here.

> to the two sides $AB$ and $CD$, while the two sides of the cut formed by the latitudinal circle correspond to the sides $AD$ and $BC$ [7, p. 32].

Seifert and Threlfall [5, p. 3] offered a visual representation for this procedure, which is shown in Fig. 3.

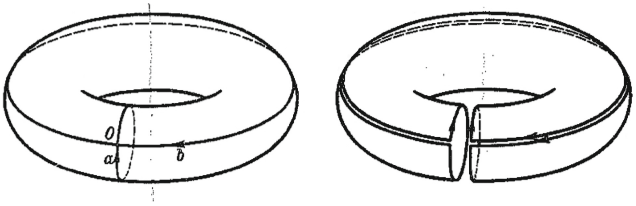


**Fig. 3.** The meridian circumference is represented by $a$ and the parallel conference by $b$. The point of intersection is labeled here H.

Is it possible to generalize this procedure to other compact surfaces, no matter if they are orientable or not? Is there a way to go from this dynamic representation to a well-defined diagram? In other words, can all ambiguities be excluded? Moreover, in this case, would such diagrams lead to an exhaustive classification of compact surfaces? Poincaré's description alone does not allow replying positively to these questions. However, three topologists working at Princeton University, James Waddell Alexander, Henri Roy Brahana and Oswald Veblen, introduced for the first time and then systematized some diagrammatic and symbolic representations that replaced it.[10]

Alexander aimed at reducing any "closed surface", i.e. a compact surface that has no boundary, to a fundamental polygon, since he believed that a precise treatment of the two-dimensional case would help solving the problem of classifying closed manifolds of higher dimensions [8]. In his words:

> In view of the difficulties that are still to be overcome before normal forms can be obtained for manifolds of more than two dimensions, a simplified treatment of the analogous problem for closed surfaces may be worth mentioning. We shall prove that all one-sided surfaces of a given connectivity $k$ are topologically equivalent and can therefore be reduced to a single normal form; also, that the same is true of all closed surfaces of a given genus $p = (k-1)/2$[11]. *Wherever the reasoning is of an intuitional character,*

---

[10] Alexander and Brahana were both Veblen's students at Princeton University. Alexander obtained his PhD in 1915, Brahana in 1920.

[11] The "connectivity" or connection order $k$ of a compact surface $S$ is related to its Euler characteristic $\chi(S)$ being $\chi(S) = 3 - k$. For the sphere $\mathbb{S}^2$, $\chi(S) = 2$ and $k = 1$. In the case of non-orientable, "one-sided" in Alexander's terms, surfaces, the $\chi(S) = 2 - g$ and thus the genus $g = k - 1$; in the case of orientable or "two-sided" surfaces, $\chi(S) = 2 - 2g$ and thus $g = (k-1)/2$ (note that Alexander uses $p$ instead of $g$).

> *it may readily be put into perfectly rigorous form by means of known theorems* ([8], p. 158, *emphasis added*).

Therefore, he proposed to reduce the normal forms to some diagrammatic representations that allow simplifying the reasoning. Orientable surfaces, after a sequence of "cuts" along closed paths and appropriate deformations, can be transformed into a simply connected surface bounded by a chain of arcs

$$a_1b_1a_1^{-1}b_1^{-1}a_2b_2a_2^{-1}b_2^{-1},\ldots,a_pb_pa_p^{-1}b_p^{-1}$$

which may be in turn deformed into a polygon. Alexander did not only reduce any closed surface to a fundamental polygon, but, most importantly, invited the reader to reason *on* and *starting from* the diagrams in order to classify closed surfaces. The diagrams are accompanied by symbolic representations in the form of a combination of letters, called "words" (see Fig. 4); the user can thus perform operations both on the polygonal forms and on the words.

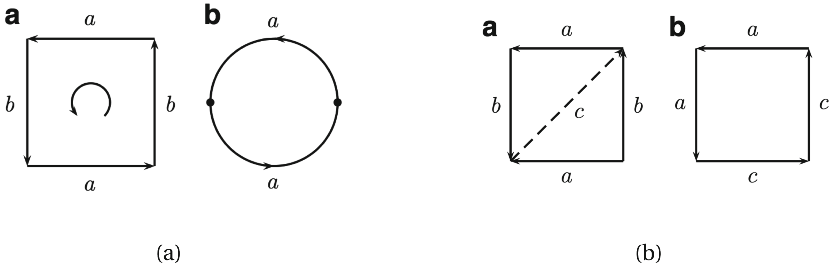


**Fig. 4.** A simplified version of the diagrams originally used by Alexander and Brahana, in [9, pp. 11–12].

In Fig. 4(a), the word *abab* corresponds to the diagrammatic representation annotated with **a** of the real projective plane. The arrows go in the same counterclockwise direction and therefore the same exponent +1 is associated to each letter. The diagrammatic representation **b** is instead similar to the one used by Dyck in Fig. 5. The (reduced) word associated to $\mathbb{P}_2(\mathbb{R})$ is *aa*. In Fig. 4(b), the word $aba^{-1}b$ is associated to the diagrammatic representation **a** of the Klein bottle. Interestingly enough, we can cut this diagram along $c$ and obtain two pieces *abc* and $bc^{-1}a^{-1}$ that can be glued back along $b$; thanks to this transformation, we obtain a second diagrammatic representation **b** of the real projective plan to which is associated the word *aacc*.

Cohn-Vossen and Hilbert showed how in the case of the Möbius band one can go from one of Dyck's *Grundform* to a polygonal form by "cuts, deformations and re-gluing" [10] (see Fig. 5). As they explain,

> We can get a model of the Möbius strip from the region between two concentric circles by identifying all the diametrically opposite pairs of points

> of the smaller circle [Dyck's *Grundform*]. At first sight it certainly would not appear that this figure represents the same surface as the square model (...). However, the square model can be obtained from [Dyck's model] by cutting the ring into two halves (Fig. 312), deforming the parts (Fig. 313), turning one half over (so as to interchange the positions of $e$ and $b'$ in Fig. 313), and finally physically re-uniting some of the pairs of edges that originally belonged together while abstractly identifying the rest (Fig. 314) ([10]. p. 317).

These representations, having a classificatory function, are thus diagrams: they are well-defined, dynamic and mathematical objects on which to calculate. We will go back to this first example in the discussion.

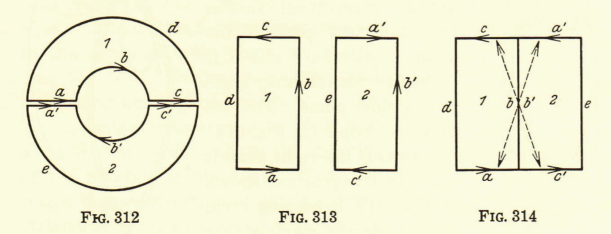


FIG. 312 FIG. 313 FIG. 314

**Fig. 5.** Figures taken from [10], p. 317.

# 3 An Example from Algebra: Complex Semisimple Lie Algebras

## 3.1 Root Systems and the Classification of Simple Lie Algebras

Our second example of the classificatory function of diagrams is taken from algebra and concerns in particular the classification of complex semisimple Lie algebras.[12] A Lie algebra $\mathfrak{g}$ over the field $\mathbb{C}$ of complex numbers is a *complex vector space* provided with a bilinear map $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ that satisfies the following properties

- $[X, X] = 0$ for all $X \in \mathfrak{g}$ (which is equivalent to $[X, Y] = -[Y, X]$ for all $X$, $Y$ of $\mathfrak{g}$),
- $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$ for all $X$, $Y$, $Z$ of $\mathfrak{g}$ (Jacobi identity).

An *ideal* $\mathfrak{h}$ of $\mathfrak{g}$ is a Lie subalgebra such that, for $Y \in \mathfrak{h}$ and for all $X \in \mathfrak{g}$, $[X, Y] \in \mathfrak{h}$. A Lie algebra $\mathfrak{g}$ is *simple* if dim $\mathfrak{g} \geq 2$ and $\mathfrak{g}$ does not contain any non trivial ideal (i.e. different from (0) and $\mathfrak{g}$). A Lie algebra $\mathfrak{g}$ of finite dimension is called *semisimple* if it does not contain any non-zero Abelian ideal.

---

[12] For an historical study on the classification of semisimple Lie algebra, see [11].

A first problem consists in characterizing complex semisimple Lie algebras. One possible characterization is to write a semisimple Lie algebra as a direct sum of simple Lie algebras. On this basis, Wilhelm Killing and Elie Cartan obtained a first classification of simple Lie algebras[13]. Note that here they did not use any diagrammatic representation or reasoning, but only work with the notion of root system that we will define below. In a nutshell, there are four families of classical simple Lie algebras:

- For $l \geq 1$, $\mathbf{A}_l = \mathfrak{sl}(l+1)$
- For $l \geq 2$, $\mathbf{B}_l = \mathfrak{so}(2l+1)$
- For $l \geq 3$, $\mathbf{C}_l = \mathfrak{sp}(2l)$
- For $l \geq 4$, $\mathbf{D}_l = \mathfrak{so}(2l)$

to which one should add five simple Lie algebras of an exceptional kind. Let $\mathfrak{g}$ be a (complex) semisimple algebra and $\mathfrak{h}$ a Cartan subalgebra of $\mathfrak{g}$, i.e. an Abelian maximal subalgebra of $\mathfrak{g}$. The dimension $n$ of $\mathfrak{h}$ is called the *rank* of $\mathfrak{g}$. A root $\alpha$ of $\mathfrak{g}$ (relatively to $\mathfrak{h}$) is a linear form on $\mathfrak{h}$ that satisfies the following property: there is a non-zero element $X \in \mathfrak{g}$ such that $[X, H] = \alpha(H)X$ for all elements $H \in \mathfrak{h}$. The set $R$ of roots of $\mathfrak{g}$ that is so obtained is called a *root system* of $\mathfrak{g}$ relatively to $\mathfrak{h}$. Such a system is independent, up to isomorphisms, of $\mathfrak{h}$. The classification of simple Lie algebras is thus reduced to a classification of such root systems.

A root system $R$ spans a vector space $V$ of dimension $n$ that can naturally be provided with a Euclidean space structure. It becomes now possible to *geometrically interpret* the system.[14] The geometric meaning of a root system becomes then explicit with the following definition: Let $V$ be a real vector space of finite dimension, with the Euclidean inner product $(\cdot, \cdot)$. We say that a finite subset $R$ of $V$ is a root system in $V$ if the following four axioms are satisfied:

(i) $R$ does not contain 0 and it spans $V$,
(ii) Let $\alpha \in R$ and $t \in \mathbb{R}$; if $t\alpha \in R$, then $t = \pm 1$,
(iii) For all $\alpha$ in $R$, the set $R$ is *stable* by reflection $\sigma_\alpha$ through the hyper-plane orthogonal to $\alpha$; in other terms, for all $\alpha$, $\beta \in R$,

$$\sigma_\alpha(\beta) = \beta - 2\frac{(\beta, \alpha)}{(\alpha, \alpha)} \in R$$

(iv) Let $\alpha$, $\beta$ be two elements of $R$ and $p(\beta)$ the projection of $\beta$ onto the line spanned by $\alpha$, then $p(\beta) = \frac{(\beta,\alpha)}{(\alpha,\alpha)}\alpha$ is an integer or half-integer multiple of $\alpha$. In other words, $2\frac{(\beta,\alpha)}{(\alpha,\alpha)} \in \mathbb{Z}$.

---

[13] See in particular, Elie Cartan's doctoral Thesis [12].

[14] For example, let $\mathfrak{sl}(3, \mathbb{C}) = \mathbf{A}_2$ be the vector space of $3 \times 3$ matrices with complex coefficients of zero trace (i.e. such that the sum of the diagonal coefficients is zero), with a bracket defined by $[A, B] = AB - BA$. This would be a Lie algebra of dimension $3^2 - 1 = 8$. The commutative subalgebra $\mathfrak{h} \subset \mathfrak{sl}(3, \mathbb{C})$, constituted by the diagonal matrices of zero trace is a Cartan subalgebra of $\mathfrak{sl}(3, \mathbb{C})$. The dimension of $\mathfrak{h}$ is $3 - 1 = 2$. The $3^2 - 3 = 6$ roots of $\mathfrak{sl}(3, \mathbb{C})$ (relatively to $\mathfrak{h}$) span a Euclidean space of dimension 2.

This definition is the starting point of Artin's and van der Waerden's investigations on simple Lie algebras around 1932–1933[15]. In the case of simple Lie algebras of rank $n = 2$, they represented each root system by a so-called "vector diagram", which satisfies the four axioms listed above.

## 3.2 Artin's Presentation

Our aim is now to present the diagrams that were used by Artin in a series of lectures given at the University of Göttingen in July 1933.[16]

In his presentation, Artin starts by *illustrating* some geometrical properties characterizing a root system[17]; then, he constructs *Vektordiagramme*, "vector diagrams", that satisfy the geometrical conditions arising from the axioms 1–4. These diagrams are thus well-defined mathematical objects; moreover, they are dynamic, because it is possible to perform transformations on them such as reflections and orthogonal projections. The user has to track these geometric transformations *on* the diagrams, in order to verify that a root system is stable for some of them, for example for all reflections on the hyper-plan orthogonal to a root, no matter which root is taken: the diagrams represent stability proprieties through their possible dynamic transformations. Artin reasons on these vector diagrams, by enumerating all the possible cases (in a two-dimensional Euclidean space)[18]. More precisely, he enumerates the diagrams corresponding to each *irreducible* root system according to the dimension $n$ of the space $V$ that $R$ generates. Let $R$ be a root system and $V$ the vector space spanned by $R$, the root system $R$ is said to be reducible in $V$ if $V$ can be written as a non trivial direct sum $V = V_1 \oplus V_2$ where $R_1 = R \cap V_1$ and $R_2 = R \cap V_2$ are root systems in $V_1$ and $V_2$. If not, $R$ is irreducible. Moreover, a system $R$ is irreducible if and only if its corresponding Lie algebra is simple. As a consequence, the classification of such diagrams leads to the classification of (complex) simple Lie algebras. From the fact that a vector diagram represents an irreducible or a reducible root system, it is possible to infer whether the corresponding Lie algebra is simple or not. For $n = 1$, the only possible diagram consists of the vectors $\pm a$ and corresponds to the simple Lie algebra $\mathfrak{sl}(2, \mathbb{C}) = \mathbf{A}_1$. For $n = 2$, Artin enumerates three possible diagrams (see Fig. 6). The diagram in Fig. 6(a) corresponds to the classical simple Lie algebra $\mathfrak{sl}(3, \mathbb{C}) = \mathbf{A}_2$, with the angle formed by vectors $\alpha$ and $\beta$ measuring $2\pi/3$. The diagram in Fig. 6(b) corresponds to the classical simple Lie algebra $\mathfrak{so}(5, \mathbb{C}) = \mathbf{B}_2 \cong \mathfrak{sp}(4, \mathbb{C}) = \mathbf{C}_2$, with the angle formed by

---

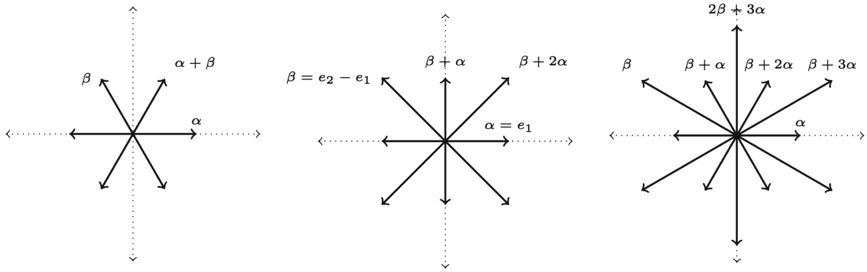
[15] See in particular van der Waerden ([13], pp. 448–449) and Artin ([14], p. 4).

[16] Emmy Noether, Hermann Weyl and Ernst Witt belonged to the audience of these lectures. A written version of Artin's presentation has been recently rediscovered by Christophe Eckes and Norbert Schappacher thanks to Ina Kersten. See their commentary on the Oberwolfach Photo Collection website, https://owpdb.mfo.de/detail?photo_id=9265.

[17] He uses the term *Veranschaulichung*, which means illustration as well as visualization.

[18] Ch. II of his lecture is entitled *Aufzählung der Diagramme* ("”enumeration of the diagrams).

vectors $\alpha$ and $\beta$ measuring $3\pi/4$.[19] The diagram in Fig. 6(c) corresponds instead to the exceptional simple Lie algebra $\mathfrak{g}_2 = \mathbf{G}_2$, with the angle formed by vectors $\alpha$ and $\beta$ measuring $5\pi/6$.



(a) (b) (c)

**Fig. 6.** Diagrams associated to simple Lie algebras of rank 2.

By contrast, the diagram in Fig. 7 represents a reducible root system $R$ of rank 2 (i.e. which spans a vector plane $V$) and is thus implicitly excluded by Artin. Indeed, in that case, $R$ can be decomposed into two systems $R_1 = \pm\alpha$ which spans a vector line $V_1$ and $R_2 = \pm\beta$ which spans a vector line $V_2$, so that the vector plane $V$ can be written as a direct sum of $V_1$ and $V_2$. As a consequence, this diagram is attached to the semisimple Lie algebra $\mathfrak{sl}_2 \oplus \mathfrak{sl}_2$ (direct sum of two simple Lie algebras).

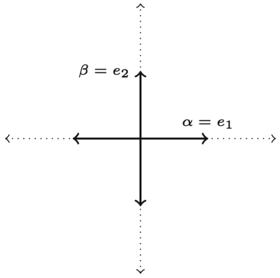


**Fig. 7.** A diagram associated to a semisimple (but not simple) Lie algebra of rank 2.

Starting from $n = 3$, it becomes very difficult to use two-dimensional representations of vector diagrams. Despite this difficulty, Artin and van der Waerden, by using these diagrams, were able to define an exhaustive classification of simple Lie algebras.

---

[19] The simple Lie algebra ($\mathfrak{so}(5, \mathbb{C})$ consists of the $5 \times 5$ antisymmetric matrices with complex coefficients.

### 3.3 Dynkin Diagrams

Starting from the vector diagrams presented in the previous section, it is possible to construct very elementary diagrammatic representations called *Dynkin diagrams*, corresponding to simple Lie algebras of an arbitrary given rank $n$. Roughly speaking, a Dinkyn diagram is a graph, whose edges can be doubled or even tripled; a simple edge is never directed, whereas a doubled or a tripled edge is. Dynkin diagrams are simplier than Artin's vector diagrams; most importantly they are two-dimensional, whatever the rank of a simple Lie algebra might be.[20]

Before arriving at Dynkin diagrams, we need to recall some mathematical definitions. Let $R$ be a root system spanning the vector space $V$. A set $\Delta$ contained in $R$ forms a *base* of $R$ if it satisfies the three following properties: (i) $\Delta$ is a base of $V$, (ii) each root $\alpha$ of $R$ can be written as a linear combination of elements of $\Delta$ with integer coefficients, (iii) these coefficients are either all non-negative or all non-positive. In the diagrams in Fig. 6, $\{\alpha, \beta\}$ represents a base for each of the investigated root systems. The elements of $\Delta$ are called simple roots of $R$ (relatively to $\Delta$).

The Cartan matrix of $R$ (relatively to a base $\Delta$) is the matrix $(n(\alpha, \beta))_{\alpha,\beta\in\Delta}$, with $n(\alpha, \beta) = 2\frac{\|\alpha\|}{\|\beta\|}\cos\varphi$, where $\varphi$ is the non oriented angle formed by vectors $\alpha$ and $\beta$. In the diagram in Fig. 6(a) corresponding to $\mathfrak{sl}(3, \mathbb{C}) = \mathbf{A}_2$, $\|\alpha\| = \|\beta\|$ and the angle formed by vectors $\alpha$ and $\beta$ measures $2\pi/3$; as a consequence $n(\alpha, \beta) = n(\beta, \alpha) = 2\cos(2\pi/3) = -1$ represent the non diagonal coefficients in the Cartan matrix $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ attached to this system. In the diagram in Fig. 6(b) corresponding to $\mathfrak{so}(5, \mathbb{C}) \cong \mathfrak{sp}(4, \mathbb{C})$, $\|\beta\| = \sqrt{2}\|\alpha\|$ and the angle formed by vectors $\alpha$ and $\beta$ measures $3\pi/4$; the Cartan matrix of this system is $\begin{bmatrix} 2 & -1 \\ -2 & 2 \end{bmatrix}$. In the diagram in Fig. 6(c) corresponding to $\mathfrak{g}_2$, $\|\beta\| = \sqrt{3}\|\alpha\|$ and the angle formed by vectors $\alpha$ and $\beta$ measures $5\pi/6$; the Cartan matrix of this system is $\begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix}$.

We can now define a Dynkin diagram of $R$ (relatively to the base $\Delta$) as a graph whose vertexes are the (simple) roots of $\Delta$. Two vertexes $\alpha$ and $\beta$ are related

- by a simple edge if $n(\alpha, \beta) = n(\beta, \alpha) = -1$,
- by a double edge if $n(\alpha, \beta) = -1$ and $n(\beta, \alpha) = -2$,
- by a triple edge if $n(\alpha, \beta) = -1$ and $n(\beta, \alpha) = -3$.

Edges are oriented from $\alpha$ to $\beta$ if $\alpha$ is strictly longer than $\beta$ i.e. if $|n(\alpha, \beta)| > |n(\beta, \alpha)|$.

---

[20] Similar diagrammatic representations were introduced by Coxeter [15], Witt [16] and finally Dynkin [17,18]. However, these diagrams do not have the same semantics and correspond to different approaches in the investigation of root systems. Despite the fact that they look similar, they have a different "dynamicity", that means one reasons differently by using them.

It is now possible to underline the classificatory function of Dynkin diagrams. First, a Dynkin diagram is *connected* if and only if $R$ is irreducible. Moreover, to each (non empty) connected Dynkin diagram corresponds *one and only one* irreducible root system (up to isomorphisms). Finally, it is possible to establish an exhaustive list of possible Dynkin diagrams, to which the classification of complex simple Lie algebras exactly corresponds (see Fig. 8). Dynkin diagrams are thus at the same time *objects of* and *instruments for* a classification. These diagrams, which were primarily introduced in order to represent in a very simple and convenient way the classification of simple Lie algebras, also appear in various other branches of mathematics and they are systematically related to classification theorems.[21]

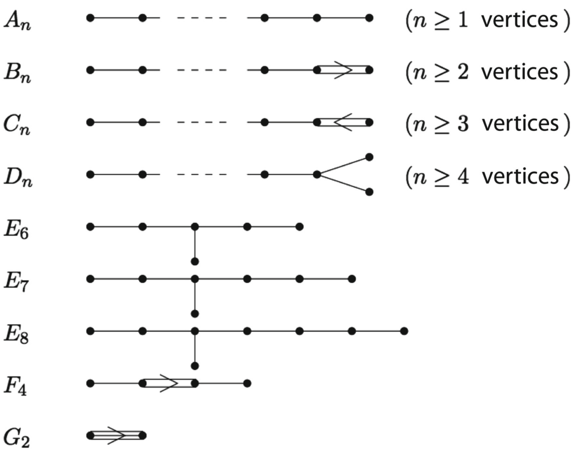


**Fig. 8.** Dynkin diagrams displaying the classification of simple Lie algebras, with the classical Lie algebras and the five exceptional ones.

## 4 Discussion

In the previous sections, we presented two examples of classifications of mathematical objects from two distinct branches of mathematics in which diagrams play a crucial role. A first result is that diagrams can have a classificatory function also for objects that, differently from those of topology, are purely algebraic, as in the case of simple Lie algebras.

The two examples have some common features. First, they both make use of diagrams, in the sense proposed by De Toffoli and Giardino [1]. Not only the diagrams but also the transitions from one diagrammatic representation to the other are well-defined, as for example the transition from Dyck diagrams to polygonal forms in the case of compact surfaces, or from vector diagrams to Dynkin's diagrams in the case of Lie algebras. However, in order to be dynamic,

---

[21] See for reference [19].

diagrams are indissolubly linked with some text. Second, diagrams allow isolating and characterizing the fundamental building blocks of the respective classifications, once ambiguities and redundancies are excluded. Third, diagrams allow conceiving the very objects belonging to the classification – think for example of the connected sum operation and of the possibility of representing it by means of diagrams, or of the Artin diagram representing a reducible root system. This is possible because diagrams are at the same time an instrument to represent some mathematical object and mathematical objects in themselves. Fourth, diagrams have a generic significance or support generality. In the case of surfaces, they allow envisioning cases that are initially problematic as generic cases – think of the real projective plane. As a whole, they amount to simplifying the forms of reasoning and the results. In the case of simple Lie algebras, diagrams allow treating uniformly the general case (classical Lie algebras) and the exceptional cases (the five simple exceptional Lie algebras), to which it would be difficult without them to associate concrete realizations *a priori*. Fifth, in both cases, we have to learn how to imagine performing some actions on the diagrams, thus developing a form of *manipulative imagination*, again in line with the case of knot diagrams (see [1]). For all these reasons diagrams, differently from illustrations such as Seifert and Threlfall's tori that are not well-defined and potentially ambiguous, appear to be particularly appropriate to serve a classificatory function. It is also interesting to point out that the fundamental polygons have become nowadays the notation that is commonly used in topology and vector diagrams are generally considered as a teaching model for understanding the meaning of Dynkin diagrams.

There are however some differences to consider between the two discussed examples. We showed that in the case of compact surfaces, the classification is discovered by means of the diagrams. In fact, the introduced representations allow understanding the case of non-orientable surfaces thus giving the classification theorem its final shape. Diagrams serve as an interface between symbolic representations such as the "words", and the procedures that one can imagine applying to some fundamental shape – a torus, the real projective plane, and so on. Such an imagined procedure, as the one described by Poincaré, is one of a kind, and remains in isolation from other mathematical representations until a well-defined diagram is introduced to replace it. However, in the case of complex simple Lie algebras, Killing and Cartan, who were the first to provide a classification, did not make use of any diagrammatic representation. In fact, vector diagrams were introduced *afterwards* to the aim of simplifying manipulations on root systems, for the reason that they display the information onto the two-dimensional space, and as a consequence in a more compact and synoptic way that allows for comparisons. Another difference is in the relation between the diagrams and the text. In the case of compact surfaces, the information on how to transform and manipulate the diagrams is present in the very representation, thanks to a careful use of arrows. However, in the case of Lie algebras, the information on how to manipulate the diagrams lies "outside" the diagrams, and for this reason much more background knowledge is needed in order to interpret the representations and understand to what their spatial features refer. This is of course related to the fact that in this case we are dealing with objects that are purely algebraic, and therefore the representations used are inevitably more abstract.

## 5 Conclusions

In light of these findings, some lessons can be drawn. First, we hope to have shown the interest in a collaboration between philosophy of mathematics and history of mathematics, for example by looking at the historical genesis of diagrammatic representations, in particular at standardization phenomena. Second, one promising strategy resides in the search of other examples that would allow proving the general impact of the classificatory function of diagrams by diversifying the envisaged domains of mathematics. Third, in general terms, more attention should be paid to the connection between the diagrammatic and the symbolic representations, and to the imagined procedures that may bring to new forms of reasoning and as a consequence to new forms of diagrammatic and symbolic representations. This is matter of current and future research.

**Acknowledgements.** This collaboration is one of the results of the project "Les mathématiques en action" (2016–2017), funded by the Archives Henri-Poincaré and the Pôle Scientifique CLCS of the Université de Lorraine. We also acknowledge support from the ANR-DFG project FFIUM for which we thank Gerhard Heinzmann.

## References

1. De Toffoli, S., Giardino, V.: Forms and roles of diagrams in Knot theory. Erkenntis **79**(3), 829–842 (2014)
2. Klein, F.: Bemerkungen über den Zusammenhang der Flächen. Math. Ann. **7**, 549–557 (1874)
3. Klein, F.: Über Riemanns Theorie der algebraischen Functionen und ihrer Integrale. Teubner, Leipzig (1882)
4. von Dyck, W.: Beiträge zur analysis situs. Math. Ann. **32**, 457–463 (1888)
5. Seifert, H., Threlfall, W.: Lehrbuch der Topologie. Teubner, Leipzig (1934)
6. Poincaré, H.: Analysis situs. J. l'École Polytech. **1**(2), 1–123 (1895)
7. Poincaré, H.: Papers on Topology: Analysis Situs and Its Five Supplements. Translated by John Stillwell. American Mathematical Society, London Mathematical Society (2010)
8. Alexander, J.W.: Normal forms for one- and two-sided surfaces. Ann. Math. **16**(1/4), 158–161 (1914–1915)
9. Gallier, J., Xu, D.: A Guide to the Classification Theorem for Compact Surfaces. Springer, Berlin (2013). https://doi.org/10.1007/978-3-642-34364-3
10. Hilbert, D., Cohn-Vossen, S.: Geometry and the imagination; transl. by P. Nemenyi. Chelsea publishing company, New York (1990)
11. Eckes, C.: (with the collaboration of Amaury Thuillier): Les groupes de Lie dans l'œuvre de Hermann Weyl. Presses universitaires de Nancy, Éditions universitaires de Lorraine (2014)

12. Cartan, E.: Sur la structure des groupes finis et continus. Doctoral thesis, Paris (1894)
13. van der Waerden, B.L.: Die Klassifikation der einfachen Lieschen Gruppen. Math. Z. **37**(1), 446–462 (1933)
14. Artin, E.: Die Aufzählung aller einfachen kontinuierlichen Gruppen. Lecture at Göttingen University, 13–15 July. The written version of this lecture is due to H. Heesch and E. Witt (1933)
15. Coxeter, H.S.M.: Regular and semi-regular polytopes. I. Math. Z. **46**, 380–407 (1940)
16. Witt, E.: Spiegelungsgruppen und Aufzählung halbeinfacher Liescher Ringe. Abh. Math. Sem. Univ. Hamburg **14**, 289–322 (1941)
17. Dynkin, E.: Classification of the simple Lie groups. Mat. Sb. **18**(60), 347–352 (1946)
18. Dynkin, E.: The structure of semi-simple Lie algebras. Uspekhi Mat. Nauk **2**(4(20)), 59–127 (1947)
19. Hazewinkel, M., Hesselink, W., Siersma, D., Veldkamp, F.D.: The ubiquity of Coxeter-Dynkin diagrams. Nieuw Archief voor Wiskd. **25**(3), 257–307 (1977)

# Mathematical Pictures

Axel Arturo Barceló Aspeitia(✉)

Instituto de Investigaciones Filosóficas, UNAM, Mexico City, Mexico
`abarcelo@filosoficas.unam.mx`

**Abstract.** There is still debate as to whether Euclidean diagrams are symbols, indexes or icons, and of what sort. I hold them to be pictorial icons that reproduce at least some visual features of their objects. This hypothesis has been directly challenged by Sherry [36] and Panza [29] among others. My aim on this paper is defending this thesis against Macbeth's [24–26] claim that if diagrams were pictures their content could not shift the way it does in Euclidean proof. To this goal I will present a broadly Gricean account of pictorial representation, where visual resemblance constraints but no fully determines reference, and then show how this account ratifies Macbeth's insights about the importance of the author's intentions in determining a diagram's content, in a way that allows for the sort of content-shifting that she has identified as key to understanding the role of diagrams in Euclidean proof.

**Keywords:** Euclidean diagrams · Icons · Representation · Resemblance Euclides

## 1 Introduction

Philosophers have long been intrigued by the many devices we have developed for mathematical practice, in particular, by the striking differences between formulas and diagrams. However, pinning down the exact difference between them has proved to be elusive, to say the least. One common-sensical way of making the distinction is arguing that diagrams are pictures while formulas are symbols. This commonsensical way of drawing the distinction can be made more precise by appealing to Peirce's [31] distinction between symbols, indices and icons. In this classification, symbols and indices roughly correspond, on Grice's distinction, to signs that have artificial and natural meaning, respectively, while icons are representations linked to their referents via "a mere community in some quality" or likeness [18]. In this classification, icons occupy an intermediate position between symbols and indices. Like symbols, they represent because of an artificial and intentional act – the act of artificially reproducing relevant aspects of their referents –, but like indices they relay on something that is naturally linked to what they depict – the features of their referents they reproduce. Paradigmatic examples of icons are figurative pictures, and other depictions [8]. However, it is not necessary for a sign to perceptually resemble its referent to be an icon. Many scientific models, for example, reproduce structural features of their target systems only, and therefore, they are icons that do not look like their referents.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 137–147, 2018.
https://doi.org/10.1007/978-3-319-91376-6_15

Euclidean diagrams are an interesting case of mathematical representations. There is still some debate as to whether they are symbols, indices or icons, and of what sort. For Giardino [16], MacBeth [24–26] and French [14], Euclidean diagrams are structural icons, i.e., they are fruitful in the discovery, understanding and proof of geometrical facts because of their homomorphism with genuine geometrical objects. In contrast, I hold them to be pictorial icons just like other paradigmatic cases of pictures, like realistic paintings or architectural models. This means that they represent their geometrical referents at least in part, because of salient visual features they share with them. This hypothesis has been directly challenged by Sherry [36] and Panza [29] among others. I have addressed some of these challenges elsewhere, and now want to focus on this paper on defending this thesis against Macbeth's [24–26] claim that Euclidean diagrams can be successfully used to prove theorems about geometrical objects they do not resemble and thus cannot be pictures of. I will argue that Macbeth's arguments misconstruct the role resemblance plays in this sort of icons – what from now on I will simply call "pictures" [3] – and thus present no challenge to my main thesis.

Before getting to Macbeth's challenge it is important to get an obvious hurdle out of the way. One might think that the very idea of geometrical pictures should be a non-starter since geometrical objects are abstract, and as such have no visual features that may be iconically reproduced. If similarity of visual or spatial features underlies the kind of resemblance that mediates pictorial iconicity, as I claim, then abstract objects cannot be pictorially represented, for they have no visual or spatial features. My response is as follows: The hypothesis I defend here does require that geometrical entities have visual features; in particular, it requires geometrical entities to be shaped in certain ways, similar to those of geometrical diagrams. I expect circles to be round, for example – not round in some sui-generis sense, but in a way that is at least similar to the way wheels and vinyl records are also round. So much is true. However, the metaphysical claim that geometrical objects cannot have sensible properties is far from a settled matter. On the contrary, it is a common thesis among historians and philosophers that the objects of Euclidean geometry are quasi-empirical [12, 16, 17, 19, 30] and that not all abstract objects and universals lack sensible features akin to those of everyday concrete objects [11, 13, 15, 23, 33, 34, 37, 38]. I will not argue for these claims here, but my forthcoming arguments ought to be read within this quasi-empirical framework.

## 2 Macbeth's Challenge

The hypothesis that Euclidean diagrams are pictures han been challenged by Macbeth [24–26] on her excellent study of Euclidean diagrams. According to Macbeth, "a drawn circle in Euclid is not usefully thought of as giving us a picture … of the thing that the word "circle" names." [26] For her, the role of diagrams in Euclidean proof largely consists in the de- and reconfiguration of content displayed by geometrical drawings, not in the analysis of a given static picture. According to Macbeth, if diagrams were pictures, no diagram representing one kind of object could be used to draw conclusions about another. A diagram of an isosceles triangle, for example, could be used to draw conclusions about isosceles triangles or about triangles in general, but not about circles or pyramids. However, argues Macbeth, there are clear cases of mathematical proof in Euclid where this is exactly what happens. Consider, for example, the proof of Euclid I.1.

In order for Euclide's proof to go through, it is essential that we regard one and the same drawn line, AB, in different stages of the proof, first as the radius of a circle, and then as the side of a triangle. Any philosophical account of the role of diagrams in Euclidean proof has to allow for content shifting of this sort.

> In a Euclidean demonstration, what is at first taken to be, say, a radius of a circle is later in the demonstration seen as a side of a triangle. But how could an icon of one thing become an icon of another? How, for example, could an icon of a radius of a circle turn into an icon of a side of a triangle? [26]

According to Macbeth, this is incompatible with the hypothesis that geometrical diagrams are pictures, since she thinks that the content of a picture cannot change in the course of reasoning about what it represents. Thus, to account for the shifting content of Euclidean diagrams, Macbeth endorses a pragmatic account where the author's intentions, as manifest in the diagram's accompanying text, play an essential role in determining its content. According to her,

> …the Euclidean diagram can mean or signify some particular sort of geometrical entity only in virtue of someone's intending that it do so and intending that that intention be recognised. One's intention in making the drawing—an intention that can be seen to be expressed in the setting out (in those cases in which there is one) and throughout the course of the *kataskeue*—is, in that case, indispensable to the diagram's playing the role it is to play in a Euclidean demonstration. [24]

Furthermore, one's intention can override what the diagram shows, so that if the geometer draws an angle with the intention "merely to draw an angle,… that which he draws… will necessarily be right, or acute, or obtuse; but [what it represents] will be neither right nor acute nor obtuse. It will simply be an angle." [24].

According to Macbeth [24], the recognition that intentions play an essential role in determining the content and role of Euclidean diagrams entails that they cannot be pictures, i.e., that even though many figures in Euclidean diagrams share some visual and spatial features with their referents, they do not represent them in virtue of this resemblance. Macbeth recognises that diagrams of circles, for example, look like circles, but argues that it is not because of this that they represent circles, but because of the interplay between pragmatic mechanisms and the structural homomorphism between the diagrams and their geometrical referents.

Thus, for Macbeth, an account of diagrammatic representation that takes seriously the importance of intentions is incompatible with the hypothesis that diagrams are pictures [26].

> Drawn figures in Euclid do not just picture various geometrical figures (any more than Arabic numerals picture collections of things); instead they display the contents of the concepts of figures in plane geometry, themselves understood in terms of relations of parts, in a mathematical tractable way. A drawn circle in Euclid is not just a picture or instance of a circle but instead an iconic display of the relation of parts that is constitutive of something being a circle. [25]

In the following, I aim to show that Macbeth is wrong in thinking that the content of pictures is fixed previously and independently of any pragmatic considerations regarding the intentions of the author. On the contrary, I will show that an adequate account of the interpretation of pictures, in general, ought to incorporate intentional concerns and, therefore, that the hypothesis that diagrams are pictures is compatible with Macbeth's recognition that intentions play an essential role in determining a diagram's content. This will allow me to show that Macbeth is wrong in assuming that pictures cannot shift content and thus, that her argument for the claim that Euclidean diagrams cannot be pictures is unsound.

## 3 A Pragmatic Account of Pictures

Resemblance guides, but does not fully determine our interpretation of pictorial icons: how a picture looks is never sufficient to determine what it represents. This means that resemblance is a necessary, but not sufficient, condition for pictorial representation. This basic insight should be enough to meet Macbeth's challenge.

Even though there is a broad debate regarding exactly what it takes for something to represent something else, there is a growing consensus in the philosophical literature that, at least in the case of what Grice called non-natural meaning – and Macbeth recognises that diagrams have non-natural meaning in this very sense –, intention and context are heavily involved [1, 2, 5, 7, 35]. Thus, in order to meet Macbeth's challenge, all I need to do is to give an account of pictorial representation as non-natural meaning and show that Euclidean diagrams are pictures in this very sense. If pictorial content is non-natural meaning, as I will argue it is, it must not come as a surprise that intention and context play an important role in the interpretation of pictures.

With this purpose in mind, I will define pictorial iconicity by extending Grice's characterization of non-natural meaning thus: a picture *p* depicts an object or state of affairs *o* iff *p* was made to look like object *o* in such a way that, under normal conditions, its audience is able to work out, on the basis of what *p* looks like, the rational assumption that it was used with a representational purpose (i.e. that it was used to represent some particular object, or objects, to a certain audience), and other background assumptions (about how the represented object looks like, about the conventions of the media, etc.) that it would be very unlikely that the user would have given it the appearance it has, resembling object *o,* unless she wanted us to recognize her intention to depict *o*.

On this account, a stick figure, for example, can be used to represent a person (to a particular audience in a given context), if it would be rational to expect from such an audience, in such a context, to figure out that, assuming that the stick figure was used with the intention of representing something and assuming certain background information both about how people look, and about what resources were available to the user (for example, how much time she had to make such a drawing), that her most likely intention in making it look roughly like a person was to represent a person.

One must be very careful in noticing that to say that a picture looks like or visually resembles its subject does not mean that looking at the picture *is just like* looking at its referent. All it means is that the picture shares enough visually salient properties with its referent for the author's referential intentions to be recognized. Consequently, when I say that a figure *f* in a diagram depicts a geometrical object *o* in a proof, all I mean is that *f* represents *o* at least partially in virtue of being drawn so as to show at least a visually salient property *p* – most commonly, its shape – that it shares with (or is sufficiently similar with the analogous property in) its referent *o*, so that the readers of the proof can be rationally expected to use this information to work out that it was drawn with the manifest intention of representing *o* (most likely, because, in the context, *o* is the most relevant object to have such shape *p,* and there is no conflicting evidence in the context about another representational intention by the author).

For this Gricean account to work, the relevant resemblance cannot be just the sharing of any visual property. It is also not enough that the relevant property be perceptually perspicuous – "syntactically salient" in Kulvicki's [21] terminology. Such conditions are too weak to trigger the Gricean mechanism. It is necessary that the common visual properties be clearly not accidental, and instead denote intentionality. In other words, it must be at least obvious to any person (in the intended audience) who looks at the figure that it was drawn intentionally so as to exhibit such properties. For example, the roundness of the figures we use to represent the circles in Euclides I.1 is salient in this sense, because it is a feature that curves do not usually have unless they are intentionally drawn to look that way. It is this manifest intentionality of the roundness of the figure that triggers the Gricean mechanism of interpretation. In other words, the audience hypothesizes that the author intended his drawing to represent a circle because that would explain her intention in making it patently round.

Besides being visually salient, the resemblance involved has to ideally strike a balance between representational costs and benefits. In the case of diagrams, it must strike a balance between how hard it would be to draw and how much it would make interpreting the diagram easier – how much it would contribute to the immediacy of its content, in Kulvicki's [21] terminology. In the context of Euclidean proof, for example, it is necessary to determine both *what* geometrical information needs to be represented and *how* best to represent it. To determine what must be represented, one must consider what is given in the initial conditions of the proof, what is to be proved, etc. To determine how to represent it, in contrast, one has to evaluate the informational and cognitive advantages and disadvantages of the available means [6]. In particular, one must decide what information is worth representing in the diagram, and what information is better left in the text. To determine which features of the geometrical situation are worth reproducing in the diagram, one must weigh both its cognitive and informational costs and benefits. Continuing with out example of Euclid's Theorem I.1., for instance, we use a roundish closed curve to represent a circle, instead of a perfect circle, because drawing it perfectly round would require too much effort without making it much easier to recognize as representing a circle. On the other hand, if the line was not closed, but open, it would lack a central feature of circles key to the proof's validity (as Macbeth has shown). Furthermore, if the line were not roundish but polygonal, the resulting diagram would be too confusing. Thus, roundness is a feature of the circle that is worth trying to reproduce in the diagram, but not perfectly.

Of course, being confusing, is not a logical defect, it is a pragmatic one. However, according to Macbeth, as long as the diagram contains the relevant information, i.e., as long as the figure is homomorphic to a circle in the relevant respects, there is nothing affecting the validity of the proof; and for this, all that is required is for it to be a closed line. Nevertheless, the closest the closed line looks to an actual circle, the easiest it is for us to identify its referent. It is not that the line looking like a circle is part of what makes it a valid means for proving stuff about circles – Macbeth is right about this – but it is part of what makes it represent a circle; and this is part of why we use round curves instead of, say, crooked closed lines to represent circles in Euclidean geometry. This is an important feature of Euclidean diagrams that is not captured by a structural account like Macbeth's.

In general, what features we decide to include in a picture will depend on the costs and benefits of including or excluding them. When we cannot include in the diagram all the information given in the setting of the problem, we have to choose which information to exclude and make sure there are enough indications, mostly in the accompanying text, as to what information is missing from the diagram. This is why the resemblance between diagrams and their referents is rarely total: Most of the times, it is not worth reproducing all the properties of the represented object in the picture; it might even be disadvantageous. Accordingly, most pictures have properties (including perceptual ones) that their referents do not have, and vice versa. This is why, for example, two-dimensional pictures can be used to represent three-dimensional objects, figures in Euclidean space can be used to represent figures in Non-Euclidean space, as long as there are other similarities and contextual clues that allow the interpreter to identify the diagram's content. I will explain in detail how context contributes to fixing the content of a diagram in the following section. This will allow me to show how a pragmatic account like mine can incorporate Macbeth's insights about the important role of manifest intentions in the interpretation of Euclidean diagrams and ultimately show that pictures can indeed shift content the way Euclidean diagrams do.

## 4 Context and Interpretation

According to the account of pictorial representation I have developed so far, how a picture looks is just part of the information the interpreter exploits in order to determine what is being represented. What a picture represents strongly depends on its context of use. As Bantinaki [5], Calderola [9], Dilworth [10], Hyman [19] and many others have insisted, visual resemblance is a many-to-many relation, i.e. different images may resemble the same object, and the same image can resemble many objects. As such, visual resemblance may restrict the kind of objects a picture can represent, but it cannot determine what it is being used to represent in every situation of use. Determining the content of a picture is not a matter of determining what it resembles the most. Extra background information is usually necessary, and depending on what background or contextual information is given, the same picture can depict one referent or another. This is why, in geometry, the same figure can be used to represent different entities or states of affairs in different contexts, even at different stages of the same proof.

To explain how context helps fix the content of pictures in their interpretation, it might prove helpful to say a little bit more about how context is exploited in human communication. For the purposes of this paper, let me adopt the well known Gricean account [18], according to which, whenever we engage in conversations, our communication is guided by a set of assumptions or maxims. These maxims include: (maxim of quality) say only what you believe to be true and of which you have enough adequate evidence; (maxim of quantity) be as informative as necessary; (maxim of relation) contribute only relevant information to the conversation; (maxim of manner) and be clear. These maxims together constitute what is known as the cooperative principle: "Make your conversational contribution such as required, at the stage at which it occurs, by the accepted purpose of the talk exchange in which you re engaged" [18]. Appealing to this maxim has proved to be helpful in explaining how we exploit contextual information to resolve ambiguities, fix extension to predicates, understand sarcasm, etc.

Assume now that our use of diagrams follows Grice's cooperative principle. In particular, assume Fig. 1 above is used in the Euclidean Proof of I.1 to communicate that *AB* and *AC* are radii of the same circle with center *A*. As drawn, points *B* and *C* stand on a closed curve that completely surrounds point *A*. This curve shows, among other salient properties, its round shape. It resembles a circle, but it also resembles (among other things, and to a lesser degree) other sorts of curves and therefore, at least in principle, could be used to represent them. Thus, it is necessary to consider which of these possible interpretations is most likely to be the one intended by the author. Without further information, the most promising hypothesis is that the diagram represents a simple figure very much like itself, i.e., a circle. Furthermore, after reading the accompanying text, we realise that this was the author's representational intention. Thus we infer that points *B* and *C* lie on the circumference of a circle entered at *A*.

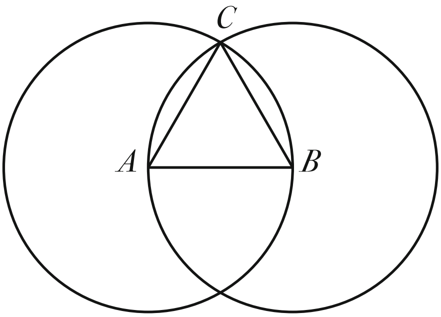


**Fig. 1.** Euclides I.1

Sometimes, however, in order to provide a consistent interpretation of the diagram that takes in consideration both what the accompanying text says and how the diagram looks, one must reject not what the diagram shows, but what the text says instead. Consider for example, Fig. 2 as used in Euclid's reductio proof in I.6 [28].

In the diagram we see two triangles sharing one side (*BC*) and one angle (*DBC*), as sated in the initial conditions of the proof. We also see that one of the triangles (*BCD*) is inside the other (*ABC*) and, consequently, is smaller. Finally, we also see that angles *ABC* and *ACB* are more or less equal. The accompanying text confirms that the angles they represent are equal. It also asks us to work under the hypothesis that *DB* = *AC*.

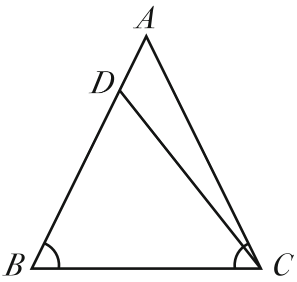


**Fig. 2.** Euclid I.6

Thus, we assume that the represented triangles *ABC* and *BCD* are of different sizes, have two equal sides (*BC* = *BC* and *DB* = *AC*) and one equal angle (*ABC* = *DBC*). However, we know from previously proved results, that if two triangles have two equal sides and one equal angle, one cannot be larger than the other, which contradicts what we see in the diagram. We have reached a contradiction. Since we need to restore consistency in order to determine the diagram's reference, and the contradiction is easily avoided if we reject the hypothesis under consideration, we do that. Once we stop trying to interpreting lines *DB* and *AC* as equal in length, we can easily identify the represented figures as two triangles *ABC* and *BCD* such that *ABC* > *BCD*, *ABC* = *DBC*, *BC* = *BC* and *AB* > *DB*. These, of course, are not impossible triangles, but regular possible triangles. This way, we can make sense of what happens in reductio proofs without having to postulate impossible geometrical objects. In general, in reductio proofs of this sort, the diagram does not represent the hypothesis to be reduced (or the contradiction reached from it), but the positive conclusion we obtain from the reductio. If a reductio proof assumes that not-*P* to get to a contradiction and thus show that P, we can expect its diagram to represent a situation where *P* holds, not one where the reduced hypothesis not-P holds, for this is impossible.

For Macbeth, "one cannot *picture* something that is impossible" [25] and, thus, the diagrams in Euclidean *reductio* proofs cannot be pictures; however, in a pictorial account like mine what is depicted in these diagrams is not the impossible state of affairs to be reduced, but the very situation that the reductio aims to prove, which is not impossible at all! Euclid's diagram for I.6 is mysterious only under the wrong impression that it is sufficient to look at a diagram to get to its content [22]. Yet, once we recognize the importance of context in determining the content of pictures, we realize that there is nothing mysterious here. Without a proper understanding of the role of contextual information in the interpretation of pictures, one might not see how a diagram can change its content, even within a single diagrammatic proof. However, once we understand that how a diagram looks underdetermines what it represents, the phenomenon of content-shifting in Euclidean diagrams is no longer surprising. Given the importance of context in helping interpreters identify the author's referential intentions, and given the importance of the author's intentions in fixing a picture's referent, it is not surprising that the content of pictures can shift the way they do in Euclidean diagrams. Thus, Macbeth's challenge poses no real threat to the thesis that geometrical diagrams are pictures.

The account of diagrams in Euclidean proof that I have presented so far takes as starting point the recognition that the representations we use to draw inferences about the world are shaped by two Gricean constraints: to include as much relevant information in the representation as possible with as little noise as possible, and to make the representation as easy to make, manipulate and interpret as possible. I have assumed, following a growing body of empirical evidence [4, 20, 27, 32] that one way to make a representation easy to interpret is by making it similar to what it represents. However, many times, making a representation resemble its referent has its costs: while making the representation easy to interpret, it can make the representation difficult to produce. It can also introduce noise, i.e., it can suggest irrelevant or misleading information. If there is relevant information that needs to be put across, but it is impractical or impossible to represent pictorially, we have to communicate it some other way. The most common way is by adding an accompanying text. Similarly, if the representation includes false information, this is also something we can fix in the accompanying text. But then, we have two different sources of information about the relevant subject; this opens the door to possible inconsistency. When inconsistencies occur, they can be resolved appealing to general pragmatic principles. They may be resolved by rejecting some of the information contained in the diagram or by rejecting some of the information contained in the text (Euclid's reduction proof of I.6 is an example). This explains why we can use the same diagram to represent different things in different contexts and why we can use a diagram in a *reductio* proof without having to postulate impossible objects.

## 5 Conclusions

In this paper I have defended a Gricean account of Euclidean diagrams as pictures. According to this account, Euclidean diagrams are pictorial icons and, as such, they exploit perceptual resemblance to fix their reference. I have also shown how this account can throw some new light on questions, like "how is it possible for the same diagram to be used in different contexts to represent different things?", "how do text and diagram interact in Euclidean proof?", and "what does the diagram in a reductio proof represent?" I have tried to show that as a consequence of the informational and cognitive constraints that shape diagrams, their visual resemblance to what they represent is usually just partial, and this results in referential underdetermination. In other words, I have tried to show that, in Euclidean diagrams, just as in pictures in general, visual resemblance constraints but does not fully determine reference. Most times, the figures in the diagram will resemble objects of more than one kind, and we will need extra information to identify the intended referent among them. Thus, it is necessary to combine the information we perceive in the diagram with information from the accompanying text. This allows for a more dynamic and malleable use of diagrams, as is manifested in Euclidean proofs where the same figure in a diagram is used to represent different geometrical objects (like Euclid's I.1) or a diagram is used to reduce a hypothesis to contradiction without actually depicting any impossible situation (as on Euclid's I.6).

**Acknowledgements.** This research was possible thanks to funds from PAPIIT IG400718 “Medio y especie: ecología y evolución desde la filosofía natural” and to helpful input from the “Tecuemepe” research group.

## References

1. Abell, C.: Canny resemblance. Philos. Rev. **118**, 183–223 (2009)
2. Abell, C.: Pictorial implicature. J. Aesthet. Art Crit. **63**, 55–66 (2005)
3. Angin, G.J., Vaez, H., Cunningham, K.L.: Visual representations and learning: the role of static and animated graphics. In: Handbook of Research for Educational Communications and Technology, pp. 865–916. Lawrence Erlbaum Associates, Mahwah (2004)
4. Arias Trejo, N., Plunkett, K.: The effects of perceptual similarity and category membership on early word-referent identification. J. Exp. Child Psychol. **105**, 63–80 (2010)
5. Bantinaki, K.: The opticality of pictorial representation. J. Aesthet. Art Crit. **66**, 183–191 (2008)
6. Barceló, A.: Pictures as epistemic tools. Sci. Stud. **14**, 45–63 (2016). https://doi.org/10.11606/S1678-31662016000100004
7. Blumson, B.: Images, intentionality and inexistence. Philos. Phenomenol. Res. **79**, 522–538 (2009)
8. Calderola, E.: Understanding resemblance in depiction: what can we learn from wittgenstein? Aisthesis **6**, 239–253 (2013)
9. Calderola, E.: Pictorial representation and abstract pictures. In: Proceedings of the European Society for Aesthetics, vol. 2, pp. 46–61 (2010)
10. Dilworth, J.: The abstractness of artworks and its implications for aesthetics. J. Aesthet. Art Crit. **66**, 341–353 (2008)
11. Deutsch, H.: The creation problem. Topoi **10**, 209–225 (1991)
12. Ferreiros, J.: Mathematical Knowledge and the Interplay of Practices. Princeton University Press, Princeton (2015)
13. Fine, K.: The problem of non-existents. Topoi **1**, 97–140 (1982)
14. French, S.: A model-theoretic account of representation (or i don’t know much about art… but i know it involves isomorphism). Philos. Sci. **70**, 1472–1483 (2003)
15. García-Ramírez, E., Mayerhofer, I.: A plea for concrete universals. Crítica **47**, 3–46 (2015)
16. Giardino, V.: Diagrammatic reasoning in mathematics. In: Magnani, L., Bertolotti, T. (eds.) Springer Handbook of Model-Based Science. SH, pp. 499–522. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-30526-4_22
17. Glas, E.: Testing the philosophy of mathematics in the history of mathematics. Stud. Hist. Philos. Sci. **20**, 115–131 (1989)
18. Grice, H.P.: Logic and conversation. In: Grice, H.P. (ed.) Studies in the Way of Words, pp. 22–40. Harvard University Press, Cambridge (1989). (1975)
19. Hyman, J.: The Objective Eye Colour, Form and Reality in the Theory of Art. The University of Chicago Press, Chicago (2006)
20. Johnson, E.K., McQueen, J., Huettig, F.: Toddlers language-mediated visual search: they need not have the words for it. Q. J. Exp. Psychol. **64**, 1672–1682 (2011)
21. Kulvicki, J.: Knowing with images: medium and message. Philos. Sci. **77**, 295–313 (2010)
22. Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. In: Chandrasekaran, B., Glasgow, J., Hari Narayan, N. (eds.) Diagrammatic Reasoning: Cognitive and Computational Perspectives, pp. 69–109. The MIT Press, Boston (1995)

23. Liebesman, D., Magidor, O.: Copredication and Property Inheritance. Philos. Issues **27**, 131–166 (2017)
24. Macbeth, D.: Realizing Reason: A Narrative of Truth and Knowing. Oxford University Press, Oxford (2014)
25. Macbeth, D.: Seeing How It Goes: Paper-and-Pencil Reasoning in Mathematical Practice. Philos. Math. **20**, 58–85 (2012)
26. Macbeth, D.: Diagrammatic reasoning in Euclids Elements. In: Van Kerkhove, D., De Vuyst, J., Van Bendegem, J.P. (eds.) Philosophical Perspectives on Mathematical Practice, pp. 235–267. College Publications, London (2010)
27. Maes, A., Arts, A., Noordman, L.: Reference management in instructive discourse. Discourse Process. **37**, 117–144 (2004)
28. Netz, R.: Greek mathematical diagrams: their use and their meaning. Learn. Math. **18**, 33–39 (1998)
29. Panza, M.: The twofold role of diagrams in Euclid's plane geometry. Synthese **186**, 55–102 (2012)
30. Parsons, T.: Nonexistent Objects. Yale University Press, New Haven (1980)
31. Peirce, C.S.: On a new list of categories. In: Proceedings of the American Academy of Arts and Sciences, vol. 7, pp. 287–298 (1867)
32. Paraboni, I., Van Deemter, K., Masthoff, J.: Generating referring expressions: making referents easy to identify. Comput. Linguist. **33**, 229–254 (2007)
33. Rosen, G.: Abstract objects. In: Salta, E. (ed.) The Stanford Encyclopedia of Philosophy (2012). http://plato.stanford.edu/archives/spr2012/entries/abstract-objects/
34. Salmon, N.: Nonexistence. Noûs **32**, 277–319 (1998)
35. Schier, F.: Deeper into Pictures. Cambridge University Press, Cambridge (1986)
36. Sherry, D.: The role of diagrams in mathematical arguments. Found. Sci. **14**, 59–74 (2009)
37. Thomasson, A.L.: Fiction and Metaphysics. Cambridge University Press, Cambridge (1999)
38. Zalta, E.: Abstract Objects. Reidel, The Netherlands (1983)

# Impact and Prevalence of Diagrammatic Supports in Mathematics Classrooms

Bryan J. Matlen[1](✉), Lindsey E. Richland[2], Ellen C. Klostermann[2], and Emily Lyons[2]

[1] WestEd, Redwood City, CA 94063, USA
bmatlen@wested.org
[2] University of Chicago, Chicago, IL 60637, USA

**Abstract.** Mathematical problem solving typically involves manipulating visual symbols (e.g., equations), and prior research suggests that those symbols serve as diagrammatic representations (e.g., Landy and Goldstone 2010). The present work examines the ways that instructional design of student engagement with these diagrammatic representations may impact student learning. We report on two studies. The first describes systematic cross-cultural differences in the ways that teachers use mathematical representations as diagrammatic supports during middle school mathematics lessons, finding that teachers in two higher achieving regions, Hong Kong, and Japan, more frequently provided multiple layers of support for engaging with these diagrams (e.g. making them visible for a longer period, using linking gestures, and drawing on familiarity in those representations), than teachers in the U.S., a lower achieving region. In Study 2, we experimentally manipulated the amount of diagrammatic support for visually presented problems in a video-based fifth-grade lesson on proportional reasoning to determine whether these multiple layers of support impact learning. Results suggest that learning was optimized when supports were used in combination. Taken together, these studies suggest that providing visual, temporal, and familiarity cues as supports for learning from a diagrammatic representation is likely to improve mathematics learning, but that administering these supports non-systematically is likely to be overall less effective.

**Keywords:** Mathematics learning · Comparison
Diagrammatic representations · Analogy · Cognitive supports

## 1 Introduction

Teaching students mathematics that is flexible, transferrable, and connected across topics is crucial to high quality instruction; however, despite decades of agreement to this pedagogical goal, many students struggle with such mathematical thinking (e.g., Polya 1954; Bransford et al. 1999; National Mathematics Panel 2008). Given the difficulties students experience in learning mathematics, a key goal is to improve instruction.

Improving mathematics instruction requires an understanding of the cognitive processes involved in acquiring mathematics knowledge and skills. In this paper, we highlight one challenge to thinking mathematically: learning to *perceive* mathematical

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 148–163, 2018.
https://doi.org/10.1007/978-3-319-91376-6_16

problems and symbolic equations as sets of relationships (arithmetic, proportional, inequality, etc.) (see Richland et al. 2012). The ability to perceive the relational structure of mathematics allows the problem solver to more easily draw connections across problems or mathematical ideas, and to think more conceptually about mathematics.

Mathematical representations are diagrammatic if they convey structural properties through their spatial and perceptual attributes. As an illustration, in the equation $3 + (4 - 5)$, a learner must know to subtract 5 from the 4 before adding three because the parentheses convey priority in carrying out the operations. Thus, part of the student's task in working with mathematical representations is to perceive the relevant structural relations that are embedded in the perceptual representations. Effective mathematical instruction, therefore, requires attention to how perceptual representations are presented in order to support students' understanding of mathematical concepts (Kellmann and Massey 2013; Richland and McDonough 2010).

Considering mathematical equations and symbolic representations as diagrams allows one to formulate insights into how to best support learners in identifying the core relationships within these representations. Diagrams can use spatial cues and sparse representations to highlight relationships rather than simply depict iconic information (Ainsworth 2006; Michal et al. 2016), however it is not the case that students learn from any and all experiences with a diagram (Rau 2017). Rather, it is clear that not only must the diagram be informative and relevant, but also pedagogical practices for supporting students' thinking and must improve the likelihood that students notice and attend to the key relationships being depicted (e.g., Richland and McDonough 2010). To improve students' attention to relationships, we can draw on strategies for ensuring that students learn from diagrams to inform mathematical pedagogy.

In the present paper, we begin by reviewing principles deriving from perceptual learning, mathematics, and reasoning literatures to highlight strategies for how to improve attention to relationships within diagrams: (a) use visual representations, make them visible while discussing them and subsequent representations, (b) use hand movements (linking gestures) to move between instructional diagrams, and (c) draw on material that is familiar to learners (see below for more detail). We then describe two sets of data suggesting that combining these strategies systematically may be the most potent way to improve student learning from mathematical diagrams. First, we describe an analysis of cross-national data collected as part of the Third International Mathematics and Science Study (TIMSS, Hiebert 2003) showing that teachers in two regions that outperform the United States, Hong Kong and Japan, used these pedagogical principles in combination more systematically than did U.S. teachers. These data suggest both that these strategies must be considered in combination, rather than as separable practices, but also that these may be correlated with student learning. In a second study, we report an experimental design in which we tested the efficacy of them being used together. The results support the correlational data identified in study one, together providing consistent indications that these are important pedagogical practices that can support students' attention during engagement with mathematical diagrams, and that doing so has consequences for student learning even when the mathematical diagram and the audio-stream of the lessons are identical.

## 1.1 Perceptual Learning in Mathematics

Though mathematics has traditionally been viewed as involving conceptual learning, there is evidence that mathematics learning is highly perceptual (e.g., Kellman and Massey 2013). For instance, mathematical concepts are frequently represented in the form of symbolic notations, which themselves contain perceptual attributes that are connected to structural properties. As an illustration, Landy and Goldstone (2010) had subjects solve simple equations that were presented either spatially consistent with the order of operations (e.g., solving $3 \ + \ (4-5))$ or spatially inconsistent (e.g., solving $3+(4 \ - \ 5))$ – these authors found that the spatial distance between influenced problem solving: Subjects were less accurate at solving simple equations when they were spaced in ways that were inconsistent with the order of operations, suggesting that even adults represent simple equations as types of diagrams.

As another illustration of perceptual learning in mathematics, interventions that support perceptual learning processes have shown some promise for improving students learning outcomes. For example, Kellman et al. have developed visual matching exercises that engage learners in linking different representations of mathematical concepts. In these exercises, students do not formally solve problems or conduct calculations; rather, students learn to identify the attributes that connect different mathematical representations. Despite never formally solving problems, students who engage in these linking activities are more accurate and at later problem solving than students that do not engage in them (Kellman et al. 2008; Kellman et al. 2010). A related perceptual learning intervention that allows students to perform physical manipulations of equations that are consistent with the grammatical rules of algebra has shown promise for supporting students' algebraic understanding (Ottmar et al. 2012).

Because mathematics involves perceiving the relevant structure in representations, diagrammatic supports that highlight structure can be a powerful tool to promote mathematical understanding and fluency (Rau et al. 2009; Rittle-Johnson et al. 2009). At the same time, simply providing diagrams may not result in successful learning (Rau 2017). Often domain leaners fail to notice relevant correspondences between representations unless highly supported in doing so (e.g., Alfieri et al. 2013; Gick and Holyoak 1980, 1983; Richland and McDonough 2010). Children and domain novices (both characteristics of k-12 school children) are most susceptible to missing key elements of comparisons and attending to irrelevant salient features that impede relational thinking, in part due to low cognitive processing resources (e.g., Richland et al. 2006). We next review research on how to support reasoning with diagrams in mathematics.

## 1.2 Diagrammatic Supports

While diagrams can serve as effective cognitive supports for learning in mathematics, students need support in comprehending diagrammatic representations before they can benefit from them. Understanding of visualizations in mathematics can be supported by cognitive aids that highlight relevant structural properties in the representations. The science of learning has made advances in understanding of how students best learn with diagrammatic representations. These principles for supporting diagrammatic fluency are discussed, below.

**Making Representations Visible Simultaneously.** Research suggests that learning in general is facilitated by the use of simultaneous diagrammatic representations (Gadgil et al. 2012; Gentner 2010; Matlen et al. 2011; Richland and McDonough 2010). In the domain of mathematics, Rittle-Johnson and Star (2007) found that middle-school age students were more likely to improve in solving algebraic equations when students compared multiple worked out equations as compared to studying them in isolation. Simultaneous presentation prompted students to compare the two domains and highlighted the relevant structural attributes of the equations. Other mathematics studies have shown learning gains when two visual representations are displayed simultaneously versus sequentially, leading to gains in procedural knowledge, flexibility, and conceptual understanding (e.g., Richland and Begolli 2015; Rittle-Johnson et al. 2009).

**Use Spatial Organization to Highlight Key Relations.** Whenever two representations are compared there are many similarities and differences that could be attended to. Learning is enhanced when the spatial organization of the representations highlights the alignments. For example, Kurtz and Gentner (2013) found participants were faster and more accurate at detecting differences in skeletal structures when two skeletal images were presented in the same orientation relative to when they were presented in a symmetrical orientation. Further, Matlen et al. (2014, in prep) found that placing images in direct spatial alignment, such that a student need not move through one object to find alignments with another, optimized the speed and accuracy with which analogies were processed. In contrast, impeded alignments were slower and led to more errors.

**Use Linking Gestures to Move Between Spatial Representations.** Linking gestures are hand movements that move between two (or more) representations that are being compared, sometimes highlighting the specific alignments between these representations, and other times simply providing support for noticing the relevance of one representation to another (Alibali and Nathan 2007, Alibali et al. 2011; Richland 2016). For instance, Richland and McDonough (2009) provided undergraduates with examples of permutation and combination problems that incorporated visual cueing, such as gesturing back and forth between problems and allowing the examples to remain in full view, versus comparisons that did not incorporate visual cueing. Students who studied the problems with visual cueing were more likely to succeed on difficult transfer problems. Linking gesture use is correlated with high mathematics learning in students (Richland 2016) and teacher gesture is well known to improve learning outcomes (see Goldin-Meadow 2003).

**Don't Overload Learners' Cognitive Resources.** Reasoning with multiple representations requires adequate working memory (WM) and executive function (EF) resources, leading to reasoning failure and lower rates of learning when resources are overloaded or non-functioning (e.g., Richland et al. 2006; Walz et al. 2000; Cho et al. 2007). When the contributions of working memory (WM) and inhibitory control (IC) were examined separately on children's successful learning and transfer from a classroom lesson based on an instructional analogy, we found that both explained distinct variance for predicting improvements in procedural knowledge, procedural flexibility, and conceptual knowledge after a 1-week delay (Begolli et al. in Press).

WM & IC were less predictive at immediate post-test, suggesting that these functions are not simply correlated with mathematics skill, but may be particularly important in the process of durable schema-formation (Begolli and Richland 2015). To reduce cognitive load during comparison, visual representations from familiar examples or domains can be used when possible in order to help students understand unfamiliar examples or domains (Duit 1991).

## 1.3 Implementing Supports in Practice

The above supports provide guidance for instructional decisions in classrooms. However, many mechanisms of learning operate simultaneously in everyday classrooms, and may augment or undermine each other, meaning that theories that explain learning in isolation may actually differ from those that explain learning in classrooms. Despite a large research base on what supports reasoning with diagrammatic representations, little research has explored the combinatorial use of supports, and this is particularly true in the context of authentic classrooms learning environments.

One reason why this may be important is that some supports may seemingly contradict one another. For instance, supports 1-3 described above are hypothesized to function because they reduce the cognitive processing load on reasoners to notice and draw inferences based on similarities between the representations; however, it is possible that adding simultaneous visual representations, spatial alignments, and gestures to process simultaneously could instead augment processing load. Thus, studying the integration of these principles is key to understanding how supports function in combination, resulting in more informed recommendations for how to best structure diagrammatic supports for classroom learning.

## 1.4 The Present Studies

The present studies examine the combinatorial use of diagrammatic supports both descriptively and experimentally. Study 1 consisted of a cross-cultural examination of the use of diagrammatic supports in middle school classrooms. The study builds on prior research by Richland et al. (2006) who coded the frequency of diagrammatic supports using a sample of eighth-grade mathematics lessons taught in the U.S., Hong Kong, and Japan from the Third International Mathematics and Science Video Study (TIMSS, Hiebert et al. 2005). These authors found that U.S. teachers regularly use comparison and contrasting cases in mathematics instruction, yet they do so without using diagrammatic supports as frequently as East Asian teachers. However, in this research, the frequency of co-occurrence of the supports was not examined. Thus, the present investigation re-analyzed lessons from the U.S., Hong Kong, and Japan to understand the combinatorial use of diagrammatic principles in these regions. In Study 2, we experimentally examine the impact of different combinations of diagrammatic supports on middle school students' mathematics learning.

## 2 Study 1

### 2.1 Method

Videodata were collected as part of the Third International Mathematics and Science Video Study (TIMSS, Hiebert et al. 2005) through a randomized probability sample of all eighth-grade mathematics lessons taught in the U.S. and seven higher achieving regions internationally. These data were analyzed and reported in a previous study (see Richland et al. 2007). The current study involved a re-analysis of codes from a set of thirty lessons that were randomly selected from the U.S., Hong Kong, and Japan. Each lesson was taught by a different teacher, and all verbalized or visually presented comparisons were identified and then coded for their presence of principles for supporting student comparison efforts.

- sourceVisAvail = the source domain of the comparison was visually available
- gestureComp = use of linking gestures for comparison
- visualAlignment = problems were spatially aligned
- sourceUnfamiliar = whether the source of the comparison was unfamiliar.

The data for the present study were re-analyzed to explore the extent to which diagrammatic supports were used in combination cross-culturally in western and eastern regions. Prior reports of this data indicated that teachers in Japan and Hong Kong were more likely to use diagrammatic supports than teachers in the U.S. (Richland et al. 2007). Thus, in the present study we combined Japanese and Hong Kong lessons in the present analysis.

The data set consisted of 588 previously coded analogies in 30 lessons from each of the three regions listed above ($n$ = 10 each). Codes for the diagrammatic supports for each analogy relative to the total number of analogies presented were averaged within each lesson.

### 2.2 Results

To determine the extent to which diagrammatic supports co-occurred with one another, Pearson correlations were conducted between the supports within each region. The results of this analysis for East Asian and U.S. lessons are presented in Figs. 1 and 2, respectively. As can be seen from Fig. 1, in East Asian lessons, supports evidenced moderate to high positive correlations (ranging from .45 to .68), indicating that supports are used moderately often in combination. In addition, all correlations within East Asian lessons were statistically significant from a zero correlation ($p$s < .05). In contrast, correlations between supports in the U.S. were inconsistent in their direction (ranging from −.48 to .29; see Fig. 2). Moreover, no correlations in U.S. lessons were statistically different from a zero correlation (all $p$s > .15). Though we view these findings as primarily descriptive, the results suggest that in addition to using diagrammatic supports less frequently in the U.S. than in East Asian countries (Richland et al. 2007), U.S. teachers are also less likely to use supports in combination.

Prior research suggested that U.S. teachers regularly use comparison and contrasting cases in mathematics instruction, yet they do so without adequately supporting

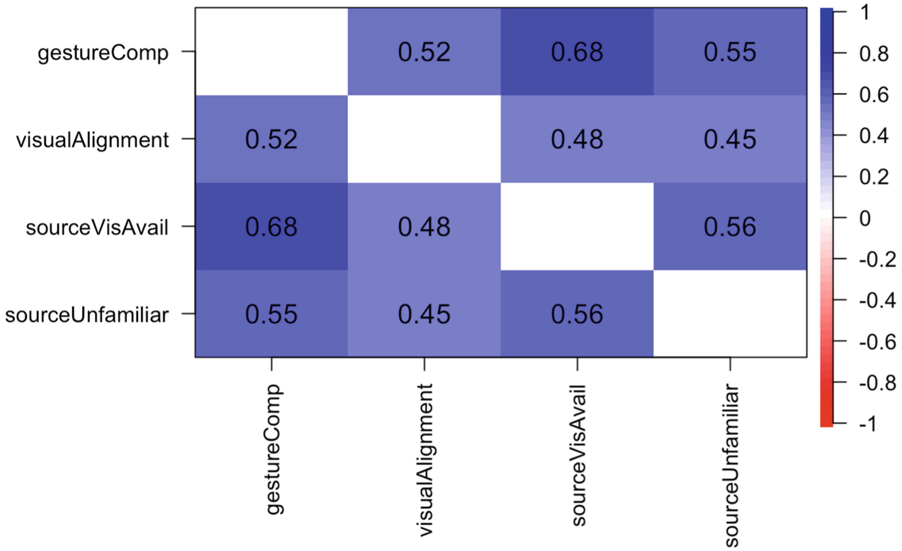


**Fig. 1.** Pearson correlations between diagrammatic supports in East Asian mathematics lessons (i.e., Hong Kong and Japan, *N* = 20). All correlations are statistically significant at alpha < .05.

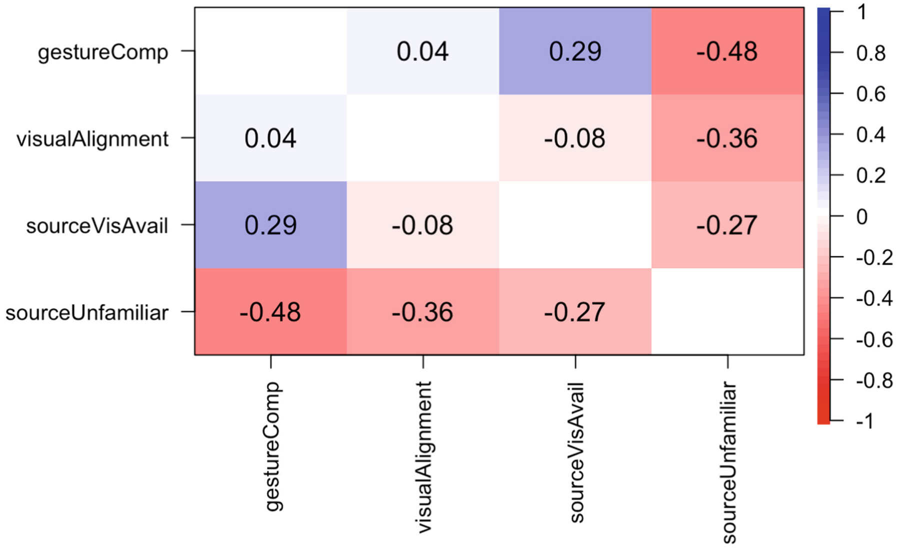


**Fig. 2.** Pearson correlations between diagrammatic supports in U.S. mathematics lessons (*N* = 10). No correlations are statistically significant (all *p*s > .15).

students in drawing these connections (Hiebert 2003; Richland et al. 2007). Importantly, teachers' rates of supporting students in drawing connections between mathematical ideas or problems during problem solving was the single factor that differentiated all higher and lower achieving peer countries in the Trends in International Mathematics and Science 1999 Video Study (Hiebert 2003). Though NCTM and disciplinary panels have long recommended helping students make mathematical connections (see National Mathematics Panel 2008; Polya 1954), this is still a serious challenge for teachers (Hiebert et al. 2005). The present study suggests that U.S. teachers also use these supports less frequently in combination than in east Asian countries – this was particularly true when the source problem was unfamiliar. This finding contrasts to East Asian teachers, who were more likely to use supports when the source problem was unfamiliar.

## 3 Study 2

Study 1 revealed that teachers in two higher achieving regions, Hong Kong and Japan, more frequently provided multiple layers of support for engaging with these diagrams in systematic ways, such that if one support strategy were used another was often used (i.e. making representations visible for a longer period, spatial alignment between diagrams, using linking gestures, and drawing on familiarity in those representations). This correlated use of support strategies mapped onto student achievement patterns, such that teachers in Hong Kong and Japan used these practices more than teachers in the U.S., which is a lower achieving region. This trend was suggestive of a relationship to achievement, but not conclusive. Thus in Study 2, we experimentally manipulated the amount of diagrammatic support provided for visually presented problems in a video-based fifth-grade lesson on proportional reasoning, to determine whether these multiple layers of support impact learning.

### 3.1 Method

The present experiment, we independently manipulated the familiarity of a source example problem with the amount of diagrammatic support provided to assess the influence of familiarity and visual supports together, separately, and in comparison to instruction without either of them. Specifically, the design was a 2 (familiarity condition: Unfamiliar or Familiar) × 2 (support condition: All Support or No Support) between subjects randomized trial.

**Participants.** Two hundred sixty-seven 5th grade students participated in this study. Forty-nine participants (18%) were excluded because they did not complete either the familiarity manipulation or one of the three math assessments. Of the remaining 218 participants, 61 were in the Familiar-All Support condition, 50 were in the Familiar-No Support condition, 56 were in the Unfamiliar-All Support condition, and 51 were in the Unfamiliar-No Support condition (See Table 1 for demographic representations of students). The study was run in nine total classrooms in five schools in the Chicago area. Four of these schools were public charter schools, while one was a Catholic school.

**Procedure.** Classrooms were visited after permission from the school's administration and teachers was granted. Each participating classroom was visited three times over a

**Table 1.** Demographics of participating students in the analytic sample.

| Demographic | Percent in sample | Number in sample total $N = 218$ |
|---|---|---|
| Females | 58% | 126 |
| African-American | 16% | 34 |
| Hispanic | 57% | 124 |
| White | 13% | 29 |
| Other race(s) | 14% | 31 |

two-week period. Students were told that the goal of the study was to understand the best ways to teach kids math.

*Visit 1.* The first visit to the classroom lasted approximately 1 h. Students completed two baseline measures, and then were randomly assigned to either a familiarity training condition or a no familiarity training condition. The baseline measures were the following:

1. *Patterns of Adaptive Learning Survey (PALS).* This 24-item measured goal orientation, and consisted of three scales (Mastery Goal Orientation, Performance-Approach Goal Orientation, and Performance-Avoid Goal Orientation) (Midgley et al. 2000). One question about students' level of math anxiety ("Math makes me feel nervous") and two questions about students' long division abilities ("I've been taught long division before" and "I can do long division") were added to the end of this survey.
2. *Content Knowledge Assessment*. This assessment was a researcher-designed test consisting of 7 items that assessed students' baseline level of knowledge of rate and ratio concepts and long division abilities (see Richland and Begolli (2015) for test properties).

On the content knowledge pre-test, students were instructed to attempt each problem and were asked to show all of their work, even if they weren't able to get a final answer. The pre-measures took approximately 45 min for students to complete.

After students completed the pre-measures, they were randomly assigned to one of the two familiarity conditions. Half of the students in each classroom were given long division instruction (Familiar condition) while the other half was given practice with long division problems (Unfamiliar condition). All students had been previously instructed in long division, these were simply opportunities to retrieve and strengthen the familiarity of these procedures. Students in both familiarity conditions were given a worksheet containing the same three long division problems. For students in the Familiarity condition, the first problem was worked out for them step-by-step, with instructions for each step. Students were asked to solve the second problem themselves, but were given those same step-by-step instructions with space next to each instruction for students to complete that step. Students were then asked to solve the third problem on their own. Students in the Control condition were given the 3 long division problems and were simply asked to solve each problem and show their work.

*Visit 2*. Researchers returned to each classroom 2–7 days later (mean = 4.3 days, median = 4.5 days), for a 90 min session. Students were assigned to one of two video-based instruction conditions: *All Supports* or *No Supports*. Assignment was random but with the constraint that one problem on the baseline test was scored and used in order to minimize any differences between baseline performance across conditions.

All of the videos contained the identical audio stream of information, and the lessons were the same teacher and classroom, but one video included more visual access to the mathematical diagrams on the board and to linking gestures, while the other video did not have these pedagogical supports.

The lesson content involved a lesson about ratio, centered on a comparison between multiple ways that different students solved a word problem involving a set of ratios. Both video lessons began with a teacher asking students to solve a ratio problem any way they would like (see Table 2). Two students in the video were then asked to share the method they used to solve the problem. The first student in the video indicated that he used the *Least Common Multiple* (LCM) method to solve the problem. The teacher then solved the problem on the board using the LCM method described by the student. A second student in the video told the class how he used division to solve the problem. The teacher solved the problem on the board using the division method described by the student. Finally, the teacher discussed the definitions of rate and ratio, summarized the lesson, and compared the two solution methods. Students in the study completed problems and answered teacher questions in a packet along with students who appeared in the video lesson.

**Table 2.** Students were provided the following prompt (accompanied by the table, below) in each video lesson: *Ken and Yoko shot several free throws in their basketball game. The result of their shooting is shown in the table. Who is better at shooting free throws?*

| Shooter | Shots made | Shots tried |
|---|---|---|
| Ken | 12 | 20 |
| Yoko | 16 | 25 |

The All Support and No Support videos differed in three ways (see Fig. 3): (1) In the All Support video, the two methods of solving the problem (LCM and division) remained visible on the board throughout the lesson. In the No Support video, the LCM solution was not visible again once the discussion of the division method began. (2) In the All Support video, these two solution methods were presented on the board in a parallel structure so that comparisons between the two solution methods could be made more easily. In the No Support video, the two solution methods were not shown at the same time, so this way of organizing the board could not support students in making comparisons between the two solution methods. (3) In the All Support video, the teacher used linking gestures while comparing the two solution methods. Gestures were not used in the No Support video. After the video lesson, students in the study completed an immediate post-test to assess how much they learned from the video lesson. Finally, students completed a 10-item survey that tested their level of engagement with the lesson.

**Fig. 3.** Screen shots of identical points in the AS and NS videos during which the teacher compared two solution strategies. Parallel organization on the board, two visible solution strategies, and gesture were used in the All Support condition, but not in the No Support condition.

*Visit 3.* Researchers returned for a third visit 4–9 days later (mean = 6.7, median = 7). Students completed 3 tasks during this visit, which lasted 1 h. As a group, students completed the d2 Test of Attention, a paper and pencil test that measures concentration and selective attention (Brickencamp and Zilmer 1998). Next, students completed the delayed post-test to assess how much they remembered from the video lesson during our second visit. Finally, students completed a demographics questionnaire.

### 3.2 Results

Outcomes of interest in the present investigation concerned students' performance on the content knowledge assessment. Specifically, we were interested in increases in the correct strategy use at post-test vs. pre-test (students' use of either the LCM or division strategies), and decreases in the incorrect subtraction strategy from pre- to post-test. For this reason, we concentrated our analyses on the problems that required students to choose a strategy and solve the problem on their own (this analysis does not include student responses to multiple choice questions). Outcome scores represent an average across all non-multiple choice questions.

**Correct Strategy Use.** To explore the presence of correct strategy use across conditions, we conducted separate 2 (familiarity condition) × 2 (support condition) between subjects ANOVAs on students' gains from pre-test to post-test (Visit 1 vs. Visit 2) and from gains from pre-test to delayed post-test (Visit 1 vs. Visit 3), using the correct use of either the LCM or division strategy as the outcome variable. The analysis for the pre- to post-test ANOVA revealed significant main effects of support condition (F(1,214) = 11.52, $p = .001$) and familiarity condition (F(1,214) = 4.36, $p = .04$), and a marginally significant interaction between support and familiarity conditions (F(1,214) = 3.11, *p* = .08) on correct strategy use. Games-Howell post-hoc tests revealed that these effects were primarily driven by the performance of students in the Familiar-All Supports condition, who performed significantly better than students in the three other conditions (*p*s < .02).

At delayed post, the ANOVA analysis revealed a significant main effect of support condition (F(1,214) = 6.73, *p* = .01), no main effect of familiarity condition (F (1,214) = .90, *p* = .35), and a marginally significant interaction between support and familiarity conditions (F(1,214) = 2.76, *p* = .10) on correct strategy use. We show delayed post performance in Fig. 4, as this time-point represents learning that is sustained over time, and is arguably the strongest test of our hypotheses. To explore the interaction, we conducted Games-Howell post-hoc tests to make comparisons between conditions. This analysis revealed that the interaction was driven primarily by higher performance in the Familiar-All Support condition relative to the Familiar-No Support condition (*p* = .02) and the Unfamiliar–No Support condition (*p* = .09).

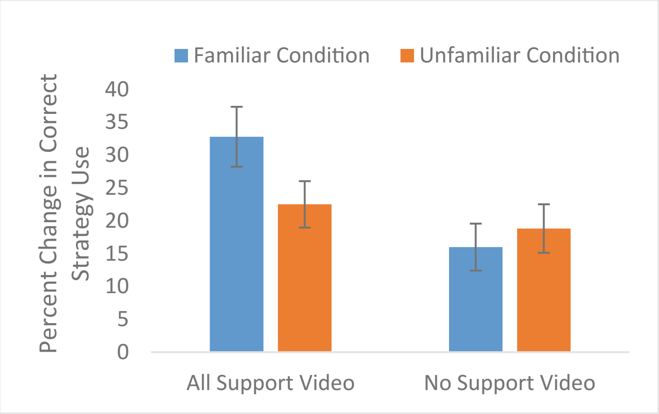


**Fig. 4.** Average percent change in correct strategy use from pre-test to delayed post-test by condition.

**Decreases in Use of The Incorrect, Subtraction Strategy.** A common incorrect strategy for comparing ratios involves the use of subtraction, where students subtract part of the whole (e.g., shots made from the total amount tried) (see Begolli and Richland 2015). To explore the use of this strategy, we conducted a 2 (familiarity condition) × 2 (support condition) between subjects ANOVA on students' decreases in use of the subtraction strategy from pre to post-test and from pre- to delayed-post-test. The ANOVA on decreases from pre to post-test revealed a main effect of support (F(1,214) = 6.35, p = .01) but no effect of familiarity condition and no interaction between support and familiarity. Post-hoc Games-Howell tests revealed a marginally significant effect for students in the Familiarity-All Support condition to decrease their use of the misconception more often than students in the Unfamiliarity-No Support condition (*p* = .06).

The ANOVA on decreases from pre to delayed post-test revealed only a significant main effect of the support condition (F(1,214) = 7.98, *p* = .005) (see Fig. 5). Marginal differences in the decreased use of the subtraction strategy were found for pair-wise comparisons between the Familiar-All Support condition vs Unfamiliar-No Support condition (*p* = .10) and the Unfamiliar-All Support condition vs the Unfamiliar-No Support condition (p = .06).

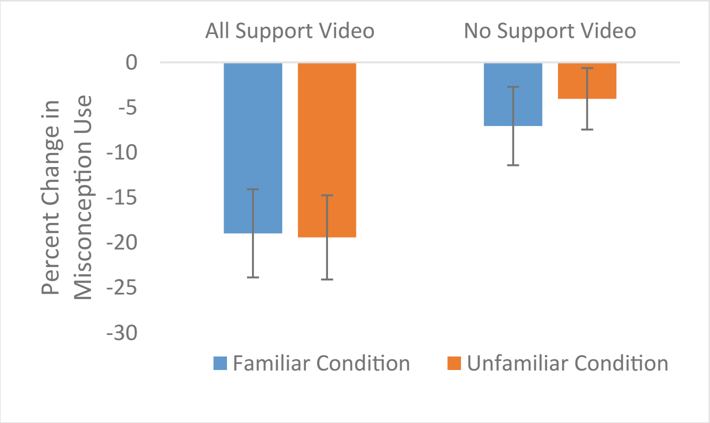


**Fig. 5.** Average percent change for the incorrect subtraction strategy from pre-test to delayed post-test by condition.

## 4 General Discussion

The studies presented here involved an observational and experimental exploration of the use of diagrammatic supports in combination in middle school mathematics classrooms. Study 1 involved a cross-cultural examination of teachers' use of diagrammatic supports and find that teachers in Hong Kong and Japan more frequently combine diagrammatic supports in mathematics lessons, whereas U.S. teachers combine supports less systematically in mathematics lessons. Study 2 manipulated the amount of diagrammatic support for visually presented problems in a fifth-grade lesson on proportional reasoning. Results suggest that learning about proportions is optimized when supports are used in combination. Though we did not explore whether combinations of supports aid learners more than single supports, these studies together suggest that providing visual, temporal, and familiarity cues as supports for learning from a diagrammatic representation is likely to improve mathematics learning.

The present research is an early attempt to explore how cognitive supports for diagrams interact in authentic environments, and is consistent with recent calls to explore instructional complexity in authentic contexts. For example, Koedinger et al. (2013) estimate that there are on the order of trillions of instructional decisions that must be made during the course of classroom teaching, and suggest that more investigations are needed that explore how instructional principles interact with one another, as well as how they interact with the content to be learned. These authors suggest that educational technology environments can used to test a large number of permutations of instructional combinations to address this problem (e.g., Koedinger et al. 2013). Similarly, we use video methodology to deeply situate this work in authentic student learning environments that are complex and that routinely combine multiple pedagogical principles, while maintaining internal validity of our experimental approach. In doing so, we attempt to understand how instructional combinations impact learning of mathematics in both an internally and ecologically valid way.

Though our approach directly examines the relationship between principle enactment and student learning, future work can more directly examine the issues that teachers confront during the course of enacting principles. For example, teachers must enact principles while they are attempting to hold both the content of the lesson and students' understanding of the content of the lesson in mind – presumably a high demand on cognitive resources – nevertheless, little work has addressed how enactment of principles influences teacher cognition. Explorations of this issue in future work might better inform theory on how principles can be optimally used in applied contexts. Moreover, our future reports will examine relationships between student characteristics known to correlate with mathematics learning, such as anxiety and executive function, to exposure to combinatorial support use and learning. This work will shed light on ways in which diagrammatic supports interact with other factors in applied contexts.

## References

Ainsworth, S.: DeFT: a conceptual framework for considering learning with multiple representations. Learn. Instr. **16**(3), 183–198 (2006)

Alfieri, L., Nokes-Malach, T.J., Schunn, C.D.: Learning through case comparisons: a meta-analytic review. Educ. Psychol. **48**(2), 87–113 (2013)

Alibali, M.W., Nathan, M.J.: Teachers' gestures as a means of scaffolding students' understanding: evidence from an early algebra lesson. In: Goldman, R., Pea, R., Barron, B. J., Derry, S. (eds.) Video Research in the Learning Sciences, pp. 349–365. Erlbaum, Mahwah (2007)

Alibali, M.W., Nathan, M.J., Fujimori, Y.: Gestures in the mathematics classroom: what's the point? In: Stein, N., Raudenbush, S. (eds.) Developmental cognitive science goes to school, pp. 219–234. Routledge, Taylor & Francis, New York (2011)

Begolli, K., Richland, L.E., Lyons, E., Wallace, E.K., Jaeggi, S., Matlen, B.J.: Executive functioning in learning mathematics by comparing representations: incorporating everyday classrooms into the science of learning. Think. Reason. (in Press)

Begolli, K.N., Richland, L.E.: Analog visibility as a double-edged sword. J. Educ. Psychol. **108**, 194–213 (2015)

Bransford, J.D., Brown, A.L., Cocking, R.R.: How People Learn: Brain, Mind, Experience, and School. National Academy Press, Washington, DC (1999)

Brickenkamp, R., Zillmer, E.: The d2 Test of Attention. Hogrefe & Huber Publishers, Seattle (1998)

Cho, S., Holyoak, K.J., Cannon, T.D.: Analogical reasoning in working memory: resources shared among relational integration, interference resolution, and maintenance. Mem. Cognit. **35**(6), 1445–1455 (2007)

Duit, R.: On the role of analogies and metaphors in learning science. Sci. Educ. **75**(6), 649–672 (1991)

Gadgil, S., Nokes-Malach, T.J., Chi, M.T.H.: Effectiveness of holistic mental model confrontation in driving conceptual change. Learn. Instr. **22**, 47–61 (2012)

Gentner, D.: Bootstrapping the mind: analogical processes and symbol systems. Cognit. Sci. **34** (5), 752–775 (2010)

Gick, M.L., Holyoak, K.J.: Analogical problem solving. Cognit. Psychol. **12**, 306–355 (1980)

Gick, M.L., Holyoak, K.J.: Schema induction and analogical transfer. Cognit. Psychol. **15**, 1–38 (1983)

Goldin-Meadow, S.: Hearing Gesture: How Our Hands Help Us Think. Harvard University Press, Cambridge (2003)

Hiebert, J.: Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study. DIANE Publishing, Collingdale (2003)

Hiebert, J., Stigler, J.W., Jacobs, J.K., Givvin, K.B., Garnier, H., Smith, M., Hollingsworth, H., Manaster, A., Wearne, D., Gallimore, R.: Mathematics teaching in the United States today (and tomorrow): results from the TIMSS 1999 video study. Educ. Eval. Policy Anal. **27**(2), 111–132 (2005)

Kellman, P.J., Massey, C., Roth, Z., Burke, T., Zucker, J., Saw, A., Aguero, K.E., Wise, J.A.: Perceptual learning and the technology of expertise: studies in fraction learning and algebra. Pragmat. Cognit. **16**(2), 356–405 (2008)

Kellman, P.J., Massey, C.M., Son, J.Y.: Perceptual learning modules in mathematics: enhancing students' pattern recognition, structure extraction, and fluency. Top. Cognit. Sci. **2**(2), 285–305 (2010)

Kellman, P.J., Massey, C.M.: Perceptual learning, cognition, and expertise. Psychol. Learn. Motiv. **58**, 117–165 (2013)

Koedinger, K.R., Booth, J.L., Klahr, D.: Instructional complexity and the science to constrain it. Science **342**, 935–937 (2013)

Kurtz, K.J., Gentner, D.: Detecting anomalous features in complex stimuli: the role of structured comparison. J. Exp. Psychol.: Appl. **19**, 219–232 (2013)

Landy, D., Goldstone, R.L.: Proximity and precedence in arithmetic. Q. J. Exp. Psychol. **63**(10), 1953–1968 (2010)

Matlen, B.J., Vosniadou, S., Jee, B., Ptouchkina, M.: Enhancing the comprehension of science text through visual analogies. In: Carlson, L., Holscher, C., Shipley, T. (eds.) Annual Conference of the Cognitive Science Society, pp. 2910–2915 (2011)

Matlen, B.J., Gentner, D., Franconeri, S.: Spatial alignment facilitates comparison (2014, in prep)

Michal, A.L., Uttal, D., Shah, P., Franconeri, S.L.: Visual routines for extracting magnitude relations. Psychon. Bull. Rev. **23**(6), 1802–1809 (2016)

Midgley, C., Maehr, M.L., Hruda, L.Z., Anderman, E., Anderman, L., Freeman, K.E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M.J., Nelson, J., Roeser, R., Urdan, T.: The Patterns of Adaptive Learning Study. University of Michigan, Ann Arbor, MI (2000)

National Mathematics Panel: Foundations for Success: The Final Report of the National Mathematics Advisory Panel, U.S. Department of Education: Washington, DC (2008)

Ottmar, E., Landy, D., Goldstone, R.: Teaching the perceptual structure of algebraic expressions: preliminary findings from the pushing symbols intervention. In: Proceedings of the Cognitive Science Society, vol. 34, no. 34, January 2012

Polya, G.: Mathematics and Plausible Reasoning: Induction and Analogy in Mathematics, vol. 1. Princeton University Press, Princeton (1954)

Rau, M.A.: Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. Educ. Psychol. Rev. **29**(4), 717–761 (2017)

Rau, M.A., Aleven, V., Rummel, N.: Intelligent tutoring systems with multiple representations and self-explanation prompts support learning of fractions. In: AIED, pp. 441–448, July 2009

Richland, L.E.: Cross-cultural differences in linking gestures during instructional analogies. Cognit. Instr. **33**, 295–321 (2016)

Richland, L.E., Stigler, J.W., Holyoak, K.J.: Teaching the conceptual structure of mathematics. Educ. Psychol. **47**(3), 189–203 (2012)

Richland, L.E., Zur, O., Holyoak, K.J.: Cognitive supports for analogies in the mathematics classroom. Science **316**, 1128–1129 (2007)

Richland, L.E., McDonough, I.M.: Learning by analogy: discriminating between potential analogs. Contemp. Educ. Psychol. **35**(1), 28–43 (2010)

Richland, L.E., Morrison, R.G., Holyoak, K.J.: Children's development of analogical reasoning: insights from scene analogy problems. J. Exp. Child Psychol. **94**(3), 249–273 (2006)

Rittle-Johnson, B., Star, J.R.: Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. J. Educ. Psychol. **99**(3), 561 (2007)

Rittle-Johnson, B., Star, J., Durkin, K.: The importance of prior knowledge when comparing examples: influences on conceptual and procedural knowledge of equation solving. J. Educ. Psychol. **101**(4), 836–852 (2009)

Waltz, J.A., Lau, A., Grewal, S.K., Holyoak, K.J.: The role of working memory in analogical mapping. Mem. Cognit. **28**, 1205–1212 (2000)

Vendetti, M., Matlen, B.J., Richland, L.E., Bunge, S.A.: Analogical reasoning in the classroom: insights from cognitive science. Mind, Brain, Educ. **9**, 100–106 (2015)

# What Sort of Information-Processing Machinery Could Ancient Geometers Have Used?

Aaron Sloman(✉)

The University of Birmingham,Birmingham, UK
a.sloman@cs.bham.ac.uk
http://www.cs.bham.ac.uk/~axs

**Abstract.** Automated geometry theorem provers start with logic-based formulations of Euclid's axioms and postulates, and often assume the Cartesian coordinate representation of geometry. That is not how the ancient mathematicians started: for them the axioms and postulates were deep *discoveries*, not *arbitrary postulates*. What sorts of reasoning machinery could the ancient mathematicians, and other intelligent species (e.g. crows and squirrels), have used for spatial reasoning? *"Diagrams in minds"* perhaps? How did natural selection produce such machinery? Which components are shared with other intelligent species? Does the machinery exist at or before birth in humans, and if not how and when does it develop? How are such machines implemented in brains? Could they be implemented as virtual machines on digital computers, and if not what human engineered "Super Turing" mechanisms could replicate what brains do? How are they specified in a genome? Turing's work on chemical morphogenesis, published shortly before he died suggested to me that he might have been considering such questions. Could deep new answers vindicate Kant's claim in 1781 that at least some mathematical knowledge is non-empirical, synthetic and necessary? Discussions of mechanisms of consciousness should include ancient mathematical diagrammatic reasoning, and related aspects of everyday intelligence, usually ignored in AI, neuroscience and most discussions of consciousness.

**Keywords:** Geometrical/topological reasoning · Evolution · Kant
Turing · AI

## 1 Introduction

Some theories of consciousness make use of mathematics, e.g. mathematical models of neural processes, but no theory that I have encountered explains how brains enable great mathematical discoveries to be made, e.g. the deep discoveries in geometry and topology, made many centuries ago, some of which, in Euclid's *Elements*, are still in regular use world-wide.[1] AI geometry theorem provers since

---

[1] A 16 page paper introducing aspects of the Turing-inspired Meta-Morphogenesis project http://goo.gl/9eN8Ks submitted to the 2018 *Diagrams* conference, was accepted as a short paper. The original version is at http://goo.gl/39DRCT.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 164–171, 2018.
https://doi.org/10.1007/978-3-319-91376-6_17

the 1960s start with logical formulations of Euclid's axioms, whereas for ancient mathematicians the axioms and postulates were not arbitrarily chosen starting points but deep *discoveries*, selected as "axioms" because other geometrical facts could be derived from them, even if originally discovered independently. Moreover, such mathematical discoveries concern necessary truths and impossibilities, which are not discoverable (or even representable) by statistics-based learning mechanisms. Necessity is not extreme probability. However, it is important not to confuse the necessity/impossibility in the *content* of mathematical discoveries with any claim that human mathematical reasoning is infallible. Many mathematicians have made mistakes that were later corrected by mathematical reasoning, sometimes triggered by empirically discovered counter examples. (However, the forms of consciousness involved in those discoveries seem to have been ignored by philosophers and scientists studying consciousness in recent decades.)

Not all geometrical reasoning is based on Euclid's axioms. Standard proofs that angles of a triangle sum to 180° use Euclid's parallel postulate, but around 1970 Mary Pardoe discovered, while teaching school mathematics, that it can be proved without using parallel lines, by considering an arrow lying on one side of the triangle then rotated in turn through each (internal) angle of the triangle. It must end up on the initial side pointing in the opposite direction, after turns totalling half a rotation, as shown in Fig. 1[2] What brain mechanisms allow such discoveries to be made and understood? As far as I know, nothing in current neuroscience or in current AI explains such discovery capabilities.

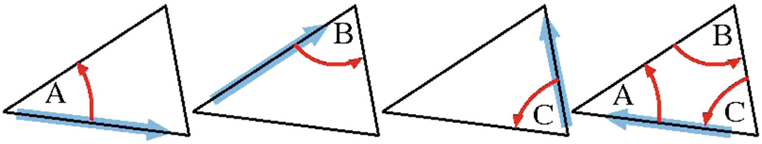


**Fig. 1.** Mary Pardoe's proof of the triangle sum theorem. Her pupils understood and remembered this more easily than the standard proof, using parallel lines. See http://www.cs.bham.ac.uk/research/projects/cogaff/misc/triangle-sum.html

Many important geometrical discoveries can be made without starting from Euclid's axioms. For example, Origami techniques allow forms of reasoning that go beyond what is provable in Euclidean geometry. Extensions of Euclidean geometry include the *Neusis* construction, known to ancient mathematicians, but not included in Euclid's *Elements*. It involves use of a movable straight edge with two marks, and allows arbitrary angles to be trisected easily.[3] The discovery of non-euclidean geometries was another important example, famously used by Einstein in his General Theory of Relativity.

Topological reasoning seems to be even more widespread, as discussed in [1]. Young children who have never studied logic or algebra can tell that it is

---

[2] See http://www.cs.bham.ac.uk/research/projects/cogaff/misc/triangle-sum.html.

[3] http://www.cs.bham.ac.uk/research/projects/cogaff/misc/trisect.html.

impossible for two linked rings made of solid, impermeable matter to become unlinked without at least one of them changing shape (e.g. ceasing to be a ring). This can be seen in their responses to clever stage magicians who make it *look* as if the impossible has been achieved. What brain mechanisms enable us to see that such things are impossible?

Some researchers seem to believe that given appropriate training, deep learning mechanisms could replicate all ancient geometrical discoveries. But statistics-based mechanisms and can only discover that certain generalisations have high, or low, probabilities. They cannot discover *necessities* and *impossibilities*, as Kant [2] showed when he pointed out gaps in Hume's classification of types of knowledge. Neural nets cannot even *express the idea* of something being impossible, or necessarily the case. Kant argued that there are important types of *non-empirical* mathematical knowledge about *necessary* truths and *impossibilities*, for which statistical evidence can never suffice.[4] What enables humans to understand these concepts, if neural nets cannot express necessity or impossibility? Is there a spatial configuration in which a planar triangle and a circle have have exactly seven boundary points in common? You can do mental experiments with imagined triangles and circles to answer this, unlike AI systems that use Hilbert's axiomatisation of geometry, and cartesian coordinates, to answer such questions, unknown to ancient mathematicians. Cartesian coordinates were not discovered until centuries later. Can any current AI system replicate that discovery?

## 2 Why Is Non-empirical Knowledge of Non-contingent Truths Important?

The kind of mathematical knowledge under discussion, is not just a philosophical oddity. It is of great practical importance to intelligent agents. Knowledge of impossibility makes it possible to rule things out without testing. Likewise knowing that having one feature of an object or process necessarily implies another allows complex decisions to be taken and used with confidence, including choice of routes in a cluttered environment, and many others.

Not only humans benefit from this kind of reasoning. Evolution has used many mathematical discoveries in selecting both physical or chemical structures and control mechanisms for those structures. Negative feedback control is used in many "homeostatic" control mechanisms from the very simplest organisms to control of blood pressure, temperature, chemical balances etc., in complex organisms. This required evolved construction kits with mathematical properties.[5]

A particular example of non-spatial mathematical intelligence in young humans is the ability to create subsuming generative grammars after many patterns of verbal communication have been found to work in the environment. This

---

[4] However, modal operators, e.g. "necessary", "impossible" should be analysed using "possible configuration" not "possible world" semantics.

[5] Some speculations about evolved construction kits are online here: http://www.cs.bham.ac.uk/research/projects/cogaff/misc/construction-kits.html.

has the great benefit of allowing novel linguistic structures to be created, or to be understood, without first learning them all from examples.

In language development, this process is followed by a further level of competence in adjusting the mechanisms to cope with exceptions to the grammatical rules. That is a rather messy kind of mathematical process. Unlike other forms of mathematical reasoning, the ability to derive new linguistic utterances to communicate novel thoughts is not guaranteed to be successful because of its dependence on the competences and vagaries of other humans.

## 3 Meta-Level Competences

A creative engineer requires additional layers of competence: meta-meta- knowledge about how to search spaces of mathematical structures to find new techniques when faced with novel problems. I don't claim that evolution produces built-in knowledge of all the kinds of mathematical knowledge used by humans: some are products of individual discovery or cooperative cultural evolution, including full understanding of cardinal numbers, which requires understanding that one-one correspondence is a transitive and symmetric relation, which Piaget's work suggests does not develop in young humans for five or six years [3].

I suspect various types of mathematical development are special cases of *staggered* gene expression: over time, brains develop new layers of meta-competence that evolved later than others, and which provide new forms of learning/discovery applicable to products of layers that evolved earlier and develop earlier in individuals, as suggested crudely in Fig. 2, allowing greater developmental leaps across generations, based on a "Meta-configured" genome. (This is very different from fashionable deep learning mechanisms.)

Individual multi-layered development seems to depend on the features of genome expression in intelligent animals summarised (roughly) in Fig. 2,[6] Recently developed genetic abstractions from previously evolved competences can be instantiated in novel ways in each generation, illustrated crudely in the figure, allowing greater developmental variety in products of a shared genome, including greater leaps across generations than could be achieved by a fixed learning mechanism provided by the genome. The history of human uses of various types of diagram seems to provide examples of this mechanism.

Current AI, including logic-based reasoning mechanisms (argued by McCarthy and Hayes to be adequate for intelligent systems [5]) and the fashionable "brain-inspired" mechanisms based on statistical learning, e.g. those surveyed by Schmidhuber in [6], cannot match the *spatial* insight-ful reasoning capabilities produced by these mechanisms. Current neural models deal with networks of nodes with numerical attributes and linked numerical relationships, whereas for the kinds of mathematical discovery I am discussing it is not necessary to collect statistical data from samples. E.g. mathematicians often reason using spatial manipulations of represented spatial structures: "diagrams in the

---

[6] goo.gl/3N1yQV gives more detail (still expanding).

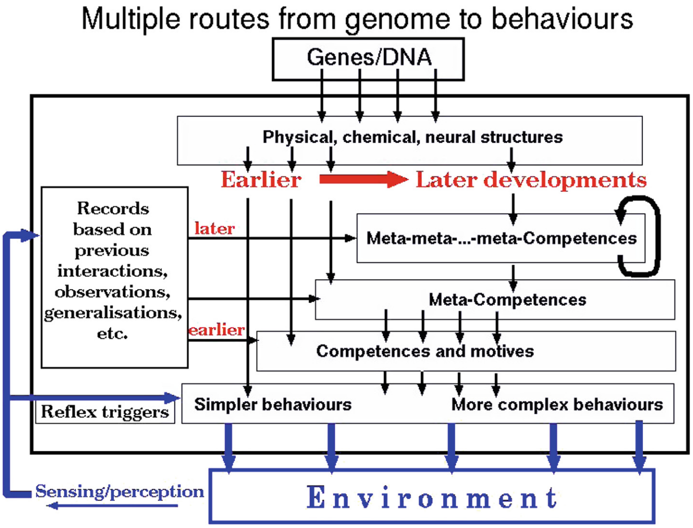


**Fig. 2.** Staggered "waves of expression" of the Meta-Configured Genome: lower layers begin development earliest via genetic influences crudely depicted on the left. Processes further to the right and higher up occur later, building on records of earlier processes that help to *instantiate* more recently evolved genetic abstractions that are expressed later in development, including new motive-generators. (Based on [4].)

mind" [7]. Perception and use of spatial affordances, by humans and other animals acting in natural environments, require abilities to perceive and reason about spatial structures and spatial relationships, including topological relationships such as containment and overlap, and partial orderings (nearer, wider, more curved, etc.), rather than precise measures [8].

## 4 Back to Ancient Mathematical Reasoning/Discovery

By examining examples of the spatial (diagrammatic) reasoning involved in ancient mathematical discoveries we may hope to gain some insights into what is missing from current forms of computation. An example that has a number of interesting features, including very easy comprehension by non-mathematicians is looking at a configuration of cup and spoon on a saucer and thinking about how to get the saucer and spoon onto the cup, using only one hand.

Similar points could be made about various stages of nest construction by birds, e.g. weaver birds,[7] that require abilities to perceive structures, select items to manipulate, moving them to new required locations, and then taking actions to enable the new items to be part of a growing stable structure.

**Conjecture:** Information processing mechanisms required for practical purposes in structured environments evolved in many species, mainly using reasoning

---

[7] Illustrated by the BBC here https://www.youtube.com/watch?v=6svAIgEnFvw.

about topological structures and relationships and partial orderings (e.g. of distance, size, speed, angle, etc.) rather than *metrical* information. In humans, the mechanisms were used in new ways, in conjunction with new meta-cognitive and meta-meta-cognitive mechanisms, leading eventually to explicit mathematical reasoning, discussion, and teaching, about topological and geometrical aspects of structures and processes in the environment.

As organisms evolve to cope with more complex structures and processes *in the environment*, they use increasingly complex abilities to create and manipulate new *internal* information structures, representing parts and relationships of external structures and processes, and supporting reasoning about consequences of possible actions, as hypothesised by Craik in 1943 [9]. Later, newly evolved meta-cognitive mechanisms, for reflecting on and comparing successes and failures of such reasoning processes, allowed new, mathematical, aspects of the structures and relationships to be discovered, thought about, and, in some cultures, communicated and used in explicit teaching and discussion. Much later, via social and cultural processes for which I suspect historical records are not available, the materials came to be organised systematically, recorded in various external "documents", such as Euclid's *Elements* and taught in specialised sub-communities.

If, as I suspect, understanding of cardinality depends on such mechanisms, then psychological evidence purporting to show innate understanding of cardinality, shows nothing of the kind: only that there are some simpler pattern recognition abilities that give observers the illusion that young children or other animals understand cardinality.

## 5 Towards a Super-Turing Geometric Reasoner

There are deep, largely unnoticed, aspects of the ways human and non-human animal minds work that are closely connected with the mechanisms underlying important non-numerical mathematical discoveries by ancient mathematicians, i.e. topological and geometrical discoveries. For ancient mathematicians the axioms and postulates in Euclidean geometry were not arbitrarily chosen starting formulae from which conclusions were derived: the axioms were all major *discoveries*, using mechanisms still available to us. And they did not use the arithmetisation of Geometry based on Cartesian coordinates.

What mechanisms allow you to discover what happens to angles of a triangle as it gets stretched by motion of one vertex relative to the other two. E.g. what will happen to planar triangle ABC, such as the triangle depicted in Fig. 3, if vertex A continually moves further from the opposite side, BC, along a line through A that intersects BC, as illustrated in Fig. 3. Even non-mathematicians can work out that as A moves further from BC the angle BAC will steadily decrease, without knowing exact lengths of lines and sizes of angles. Despite being so obvious to non-mathematicians, this answer has surprising mathematical sophistication. It involves both the continuum of locations of the angle A, and the continuum of sizes for the angle A, and a systematic relationship between the

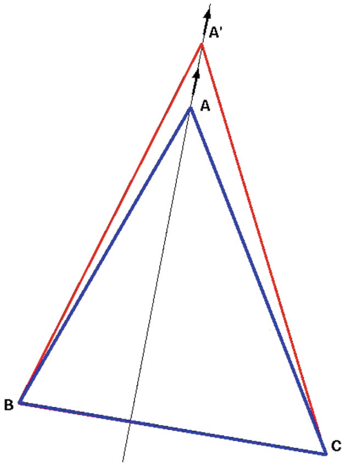


**Fig. 3.** How does the angle at A change as A moves further from BC, along a straight line that passes between B and C, e.g. moving to A′? What brain mechanisms allow reasoning about such questions?

two continua: as the distance increases the angle size decreases. It is not obvious *exactly* how the angle size and the length are related, though it is obvious that as one increases the other decreases, unless the line along which A moves intersects the line through B and C outside the segment BC.[8] What mechanisms would enable a future robot to find the relationships in Fig. 3 as obvious as we do? Brains seem to do much that is not explained by current neural net mechanisms nor by current AI models of spatial reasoning using logic or logic plus algebra, trigonometry etc. How might the mechanisms differ from a Turing machine,with its linearly ordered tape, divided into locations each of which can contain exactly one symbol?

## 6 Conclusion

A challenging research problem is to find a way to specify a type of machine that could replace a Turing machine's tape, tape-head, and symbol table, with something like a membrane on which marks can be made and which can be stretched, rotated, translated, and its new position compared with the old position, to see what has changed, with at least two layers of meta-cognition detecting and reasoning about what does and does not, and what can and cannot change. Humans thinking about the triangle problem seem to construct imagined states

---

[8] The case where A moves along a line that intersects BC outside the triangle is discussed in another document. See http://www.cs.bham.ac.uk/research/projects/cogaff/misc/apollonius.html. Surprising additional complexities are discussed in that and http://www.cs.bham.ac.uk/research/projects/cogaff/misc/deform-triangle.html.

on which very much more complex operations can be performed, including two or more co-ordinated continuous changes, and two or more levels of meta-cognition operating in parallel: e.g. one detecting and summarising changes, and another reasoning about the nature of those changes – e.g. discovering necessities and impossibilities. Is there a minimal set of basic mechanisms (perhaps chemical mechanisms in brains?) from which all the forms of spatial reasoning required for an intelligent animal can be derived?[9]

The Meta-Configured genome hypothesis sketched above implies that intelligent animals do not have a uniform innate learning mechanism that operates from birth on increasingly complex and varied data sets. Different mechanisms, with different evolutionary origins modified to fit the individual's environment, come into operation at different stages of development over an extended time period. Compare language development and Karmilof-Smith's ideas about "Representational Re-description" [10].

I suspect Alan Turing may have been working on a problem of this sort when he wrote "The chemical basis of morphogenesis" [11], now his most cited paper. What would he have done if he had not died two years after it was published?

## References

1. Sauvy, J., Sauvy, S.: The Child's Discovery of Space: From Hopscotch to Mazes an Introduction to Intuitive Topology. Penguin Education, Harmondsworth (1974). Translated from the French by Pam Wells
2. Kant, I.: Critique of Pure Reason. Macmillan, London (1781). Translated (1929) by Norman Kemp Smith
3. Piaget, J.: The Child's Conception of Number. Routledge & Kegan Paul, London (1952)
4. Chappell, J., Sloman, A.: Natural and artificial meta-configured altricial information-processing systems. Int. J. Unconv. Comput. **3**(3), 211–239 (2007)
5. McCarthy, J., Hayes, P.: Some philosophical problems from the standpoint of AI. In: Meltzer, B., Michie, D., (eds.) Machine Intelligence 4, pp. 463–502. Edinburgh University Press, Edinburgh (1969). http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html
6. Schmidhuber, J.: Deep learning in neural networks: an overview. Technical report IDSIA-03-14 (2014)
7. Sloman, A.: Diagrams in the mind. In: Anderson, M., Meyer, B., Olivier, P. (eds.) Diagrammatic Representation and Reasoning. Springer, London (2002). https://doi.org/10.1007/978-1-4471-0109-3_1
8. Sloman, A.: Predicting affordance changes (alternative ways to deal with uncertainty). Technical report COSY-DP-0702, School of Computer Science, University of Birmingham, Birmingham, UK, November 2007. goo.gl/poNS4J
9. Craik, K.: The Nature of Explanation. CUP, London (1943)
10. Karmiloff-Smith, A.: Beyond Modularity: A Developmental Perspective on Cognitive Science. MIT Press, Cambridge (1992)
11. Turing, A.M.: The chemical basis of morphogenesis. Phil. Trans. R. Soc. Lond. B **237**(237), 37–72 (1952)

---

[9] Later developments of the idea of a Super-Turing machine will be added here: http://www.cs.bham.ac.uk/research/projects/cogaff/misc/super-turing-geom.html.

# Interpreting Diagrammatic Reasoning – Between Empiricism and Realism

Michał Sochański(✉)

Poznań, Poland
`michal_sochanski@poczta.onet.pl`

**Abstract.** Diagrams are frequently used in mathematics, not only in geometry but also in many other branches such as analysis or graph theory. However, the distinctive cognitive and methodological characteristics of mathematical practice with diagrams, as well as mathematical knowledge acquired using diagrams, raise some philosophical issues – in particular, issues that relate to the empiricism-realism debate in the philosophy of mathematics. On the one hand, it has namely been argued that some aspects of diagrammatic reasoning are at odds with the often assumed *a priori* nature of mathematical knowledge and with other aspects of the realist position in philosophy of mathematics. On the other hand, one can claim that diagrammatic reasoning is consistent with the realist epistemology of mathematics. Both approaches will be analyzed, referring to the use of diagrams in geometry as well as in other branches of mathematics.

**Keywords:** Philosophy of mathematics · Diagrammatic reasoning Epistemology

## 1 Introduction

Diagrams have been used in mathematics since its very beginnings and are still a crucial aspect of mathematical practice. Geometry is the branch of mathematics that is naturally associated with the use of diagrams, however, various diagrammatic representations are also used in almost all mathematics, including, e.g., analysis, graph theory, topology or even algebra. Moreover, use of diagrams and more generally visualization has increased in recent years due to the possibilities offered by computers. Diagrams have a reputation of being effective heuristic tools, greatly facilitating our intuition and understanding of mathematical objects, as well as taking part in such aspects of mathematical practice as discovery, explanation or reasoning. Many characteristics of diagrammatic representation which are opposed to those of sentential representation (in general – not only in mathematics) have been pointed at, as well as many ways to define the notion of the diagram itself[1]. It is beyond the scope of this text

---

M. Sochański—Independent scholar.

---

[1] Properties of diagrams as opposed to sentential representation have been analyzed, among others, in [10, 11].

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 172–179, 2018.
https://doi.org/10.1007/978-3-319-91376-6_18

to analyze those differences in depth, its primary aim being to investigate how the differences between diagrammatic and sentential representations affect certain aspects of our philosophical view of mathematics. In particular, I will consider to which extent certain aspects of diagrammatic reasoning, as well as of other uses of diagrams and visualization in mathematics, have relevance to the empiricism-realism debate in the philosophy of mathematics. Some characteristics of diagrams, e.g., the role of perceptual experience in gaining knowledge about them or experimentation with diagrams might seem to be inconsistent with the *a priori* character of mathematics and to suggest that diagrammatic reasoning has at least partly empirical character. In my paper I will firstly try to elucidate reasons for such claims and secondly to consider if the empirical, or *quasi*-empirical aspects of diagrammatic reasoning pose a problem for the realist and how they fit into the realist epistemology of mathematics in general.

Before we go on, some initial remarks about realism and empiricism in mathematics have to be made. Even though there are many versions of both the realist and empiricist standpoints, it is possible to list the following main assumptions of realism, each of which is typically rejected by a *quasi*-empiricist[2]:

R1. The subject-matter of mathematics consists of mathematical objects and/or structures which exist independently of the mathematician. In the radical, Platonic form of realism it is claimed that this subject-matter is immaterial and eternal.
R2. Mathematical truth is objective – mathematical propositions have truth values independently of human cognition and language.
R3. Mathematical propositions are not only precise and certain but also necessary and in consequence – irrefutable.
R4. Mathematics is fundamentally different from the other sciences. The apriority and necessity of their propositions as well as ideal character of their subject-matter are in sharp contrast with the character and objects of empirical knowledge.

The first characteristic of realism is the strongest and is rejected or modified by some philosophers, R2–R4 are typically defended within most versions of realism[3]. Philosophers that sympathize with any form of empiricism typically reject the above claims. And so they would claim, that: all mathematical cognition stems from sensual experience, that mathematical truth is relative to language and perhaps even culture, that there is no such thing as "absolute" necessity or apriority (mathematical knowledge changes over time) and finally – that mathematics is not fundamentally different from empirical sciences but rather the difference between them is a matter of degree. Those

---

[2] It has to be noted that realism is typically opposed to anti-realism, in that the latter claims that mathematical objects do not exist. The dichotomy empiricism-rationalism, on the other hand, concerns the source and nature of mathematical knowledge. However, every realist is an apriorist in that he claims that we come to know mathematical objects independently of physical experience. In that sense it is possible to oppose realism and empiricism.

[3] It should be noted that R1-R3 are independent of each other (although clearly strongly related) and so any combination of them can be held within one position.

will be the two opposing views about mathematics, constituting the philosophical background against which the nature of diagrammatic reasoning will be discussed.[4]

In the remaining part of the paper I will first briefly discuss some empirical aspects of geometry, turning to the *quasi*-empirical characteristics of diagrammatic reasoning in mathematics in general. In the last Sect. 1 will try to point at problematic issues related with giving a positive account of diagrammatic reasoning within realist philosophy of mathematics.

## 2 Diagrams in Geometry

Geometry can be – and has been – claimed to have empirical character in a global and local sense. In the first sense, the whole of geometry can be attributed empirical character, in the second one only particular reasoning types. Whereas the empiricist will naturally locate the source of geometrical knowledge in sense experience, the realist has to explain to what extent the sensual experience takes part in learning of geometric concepts and in what sense geometry grants us with *a priori* knowledge, making use of spatial representations which are the geometric diagrams[5]. According to Plato, the proponent of the most radical version of realism, immaterial mathematical objects are accessed by a faculty called *dianoia*, which however is indirect in the sense of being mediated by the use of mathematical concepts and representations. Those representations are not the actual subject matter of mathematics, their physical form drawing us away from the true, eternal subject matter of mathematics[6], being, however, indispensable in mathematical practice[7]. On the other hand, many apriorists of the modern period, like Descartes, held that the subject matter of geometry is physical extension, which at the same time provides us with certain and general knowledge.

In the remaining part of this Sect. 1 will take a closer look at selected reasoning types in geometry, and consider whether one can raise objections as to their *a priori* character. As it is known, geometry became fully formalised and it is no longer possible to claim that diagrams are indispensable for formulation of any geometrical proof (as Kant has claimed) and thus that any geometrical proposition is *a posteriori* in some absolute sense. However, one can still ask if particular reasoning types, as they have

---

[4] I will intentionally omit the two other big families of views on mathematics – varieties of neo-Kantian positions and logical positivism, in order to focus my attention on the empiricism-realism debate. Both claim that mathematics is *a priori*, but just as the followers of Kant hold that the apriority rests on in-born structure of our minds, the followers of logical positivism defend the view that they rest on the analyticity of mathematics.

[5] As far as ontology is concerned, it is has been claimed that the subject-matter of geometry is extension, empirical space or shape. John Stuart Mill, in turn, argued that the subject matter of geometry are the actual physical diagrams.

[6] "These things themselves that they mold and draw, of which there are shadows and images in water, they now use as images, seeking to see those things themselves, that one can see in no other way than with thought" [7].

[7] Plato names five aspects of a circle: "The first is the name, the, second the definition, the third. the image, and the fourth the (…) knowledge, intelligence and right opinion" [8], stressing that each of them is a necessary step in forming our knowledge of the true geometrical objects.

been conducted before the XIX century and are still conducted now, can be claimed to be *a posteriori*. Let us assume the most commonly accepted definition of the *a pr*iori, relating it to reasoning types[8]:

> A reasoning is *a priori* iff it is justified independently of experience.

Assessing, whether a particular reasoning is *a priori* will fall into two stages: firstly, it has to be determined which aspect of the reasoning can be classified as "experience" or referring to experience. Secondly, one has to ask, whether the experience was in fact used in the justification, and if so – what role exactly it played.

Geometry abounds with arguments, in which certain visual properties of diagrams, that is their physical characteristics like shape, location of certain points, etc. are used during the reasoning process by reading them off the diagram[9]. In this case, we can understand experience to be observation of the diagram and noticing the appropriate relations between its parts. Many authors have stressed that experience is used in such cases in a different way than "usual" sense experience is used when analysing empirical objects. First of all, we are not using it as inductive support of any claim about physical objects. Secondly, the subject of geometric arguments is not the physical diagram itself, but rather what we take it to be using our precise geometrical concepts. This point is stressed by Annalisa Coliva, who claims that proofs in Euclidean geometry are *a priori*, as "the justification for holding a given theorem they provide us with depends merely on the geometrical concepts involved" [2][10].

Second type of argument I would like to mention involves using imagery to perform dynamic visualisation. A famous example of such reasoning is called reasoning by superposition and can be found in the proof of theorem I.4. in Euclid's *Elements*. Within it, the reasoner is asked to apply one triangle to another by moving it in visual imagination, the possibility of performing the movement being used as a crucial element in the reasoning. The experience used in this case seems to be of fundamentally different type than in the previously analysed case. We are not only reading topological characteristics off the diagram, but using visual imagery in moving geometric objects in our inner visual space. This has raised doubts of some philosophers, especially those who claim that mathematical objects are static and should not be associated with motion. A sixteenth century philosopher Jacques Peletier for example raises the objection that "the method of superposition does not suit the dignity of geometry because it has something mechanical about it" [5]. Although proofs by superposition

---

[8] By a "reasoning" I will understand the whole psychological process that starts with assumptions and a given diagram (or diagrams) and results in appearing of a belief state which is the conclusion of the reasoning.

[9] Visual characteristics typically fall into two categories: topological and metric. Metric properties refer, e.g., to the length of lines or size of angles. Topological properties in turn, include betweenness, incidence and inclusion. It has been claimed that only topological properties have been used in Euclid's *Elements*.

[10] According to Coliva we can distinguish between perceptual concepts which correspond to deliverances of the senses and geometrical concepts which are of purely mathematical nature and for example used when we take a straight line that is not perfectly drawn, to represent a straight line. A similar distinction has been made and analyzed by Marcus Giaquinto who has also argues that geometrical arguments that use diagrams in a significant way are *a priori* [3].

are now a historical peculiarity, the role of motion in mathematics can still be said to raise epistemological issues[11].

In conclusion, let me repeat that the term "experience" can be differently understood when related to geometrical arguments. If observation and mechanical visual imagery are to be seen as experience, then some Euclidean arguments make use of experience. However, they are not used as usual perception in getting to know about physical objects and they are always used in *Elements* in a way that avoids proving false statements. However, it can be noted that the character of the experience may matter in judging the value of a particular reasoning from a given philosophical standpoint – in particular mechanical intuitions can be claimed to be suspect from the Platonic point of view.

## 3 Diagrams in Mathematics and *Quasi*-empiricism

The above mentioned definition of *a priori* refers strictly to the role of justification and has so far been discussed only in relation to geometry. However, as it is well known, diagrams are mostly used in mathematical practice in other contexts: they are excellent heuristic tools often used to discover new facts, to explain or simply to broaden our understanding of mathematical concepts. Considering a broader context of this large variety of epistemic functions of diagrams will allow us to take into account diagrams in other branches of mathematics which do not share geometry's proximity to sensual experience. Those aspects of mathematical practice will not directly affect the theses about the *a priori* nature of mathematics, but can be viewed as *quasi*-empirical. In particular, they can be claimed to share some characteristics with the methodology of empirical sciences.

Before we go on, some clarifications have to made. The first point is that diagrams, as applied in most other branches of mathematics, do not have to be related in any way to space, shape, extension or any other object of the empirical world.[12] Instead, they are used as a possible representation of mathematical objects, that has its specific characteristics, advantages and disadvantages, hinted at in the introduction to this paper. In representing a mathematical object by a diagram we specify semantic rules by *choosing* specific shapes, dots, colours, etc. to represent mathematical objects such as relations, groups or numbers, which by themselves need not have spatial character.

In such case what we will call diagrammatic reasoning will typically consist of three stages. The first is translation of mathematical concepts into diagrammatic ones (shapes, lines, etc.). In the second stage we observe the diagram, manipulate it or experiment with it. In the third stage, we translate the effect of those activities back to sentential language. The fact of using diagrams in this way by itself does not of course

---

[11] Mark Greaves notes for example, that according to Helmholtz, "kinematic properties are not given *a priori*, but rather are derived empirically from the subject's own perceptions of bodily movements, and thus that there is a necessary linkage between the truths of geometry and the laws of mechanics"[4].

[12] An exception could be e.g. topology, which studies properties of shape in a more general and abstract way that geometry.

have to be interpreted as having any consequences to the empiricism-realism debate. However, different methods used in the second stage of diagrammatic reasoning may be considered to have *quasi*-empirical character in the sense of being similar to the study of empirical objects rather than to deductive methods typical of mathematical practice. *Quasi*-empirical methods include, in particular, “visual” experimentation, understood to be any manipulations of diagrams or creation of multiple diagrams in order to gain knowledge about them, which activities are currently mainly performed on computers. Such experiments often include gathering data and observation, as well as using induction to confirm hypotheses, thus resembling the study of an object existing independently of the reasoner. Similarly as with experiments in the empirical sciences, we also often do not know “in advance” what the outcome of the calculation will be. In this sense, methodology of experiments can be called *quasi*-empirical.

## 4 Towards a Positive Realist Account of Diagrammatic Reasoning

In the previous sections it was discussed whether diagrammatic reasoning in mathematics can be said to have some empirical or *quasi*-empirical characteristics. In the last section we will consider how a positive account of diagrammatic reasoning within a realist position may be formulated. Should the realist give a separate account of knowledge gained with use of diagrams or does it not differ from other types of mathematical knowledge? Does she face any problems when trying to formulate such an account?

Answering those questions, let us first remind that realism, in claiming that mathematics is *a priori*, postulates that we have access to its objects other than just sense perception. Realists such as Kurt Gödel have claimed that this access can be gained with the use of intuition, understood as a special cognitive power enabling us to “see” truths about immaterial mathematical objects. One possible consequence of such a view seems to be that choice of representation type should not at all be significant for the realist. If mathematical objects exist independently of our cognition and in particular – of our language, mathematicians should be able to choose whichever notation suits them[13]. Diagrams would then simply provide us with another type of access to the realm of mathematical objects. Some realistically inclined philosophers have indeed claimed that some diagrams (especially well-designed computer visualisations) allow them to glimpse into the Platonic world of mathematical objects. However, it remains unclear how the realist should explain the role of perception in diagrammatic reasoning and the way in which diagrams refer to abstract mathematical objects. Gödel has formulated a well-known analogy between perception and mathematical intuition, which however only provides a metaphor that is intended to help us understand how we come to know mathematical objects, but does not explain the role of actual perception in the analysis of visualisation. One possible answer could be that the diagram is in some way similar (or at least homomorphic) to the actual mathematical object. Such

---

[13] In a similar vein, realism allows the use of Axiom of Choice or impredicative definitions.

interpretation may initially seem appealing, however, on closer look, one can doubt how it is possible to state anything about similarity of a physical object, a man-made representation, with the ideal object of mathematics. In fact, radical realists such as Plato himself, would not at all be ready to admit such similarity. In addition to that, it is hardly possible to claim that diagrams point to mathematical objects "by themselves". One diagram can represent many objects, depending on how the semantic rules translating the physical characteristics into mathematical properties are formulated[14]. An account of the role of observation has been given by James Robert Brown, according to whom diagrams can grant us access to the objects of mathematics – "as telescopes help the unaided eye, so some diagrams are instruments (rather than representations) which help the unaided mind's eye" [1]. Brown does not, however, explain how diagrams allow us to learn about non-material mathematical objects, suggesting, however, that perception of a diagram induces mathematical intuition or perception of mathematical objects.

Realism also faces problems when explaining the nature of mathematical experiments. On the one hand, one can interpret them in a realist setting as giving us insight into the Platonic world of mathematical forms. In fact, when generating the Mandelbrot set we can have a sense of facing something external to us which has properties that are objective. However, it does not seem that we are experimenting with the actual mathematical objects – in whichever way we understand them. Instead, we seem to be experimenting with the mathematical notation itself, finding out which consequences will follow from adopting specific semantic rules in creation of a visualization. Finally, it seems that realism should account for the role of spatial intuition in mathematics, specifying how it relates to perception on the one hand and to mathematical intuition on the other. Spatial intuition, outside of the realist setting, is usually considered to be a cognitive faculty that allows us to consider spatial representation of mathematical objects, employing the senses in a constructive way in order to analyse precise mathematical concepts.[15] However, most realists, including Plato or Gödel, only refer to perception and intuition that connects us "directly" with non-material objects[16]. It seems that an account of spatial intuition could explain how perceptual content links with the objective subject-matter of mathematics.

In conclusion, it seems that *quasi*-empirical aspects of the use diagrams and visualisation in mathematics can be interpreted within both the empiricist and realist positions. The *quasi*-empiricist might claim that they only show how the boundary between mathematics and the empirical sciences is not sharp. Mathematical realist, in turn, can hold that diagrams are just one way of accessing the objective realm of

---

[14] This point has been made by David Sherry, who claims that the Platonist cannot explain "use of the same diagram in proving theorems about radically incompatible figures" [9].

[15] To give just one example, Felix Klein used the notion of "refined intuition" which made use of sensual experience, but was "armed' with precise mathematical concepts.

[16] One exception was Proclus, who adds another cognitive faculty to the ones postulated by Plato, which he calls imagination (*phantasia*). In words of Dmitri Nikulin, imagination is "intermediate between sense perception and discursive reason: with the former, imagination shares the capacity to represent geometrical figures as extended; with the latter, it shares the capacity to represent its object as unchangeable according to its properties" [6].

mathematical objects. An analogy with perception of physical objects can be evoked here, in that visualisation allows us partial glimpses into that realm, giving the mathematician an impression of looking at it from a distance. It is, however, not clear, if such impressions should be treated only as metaphors or if they are to be attributed to particular semantic properties of diagrams as representation types, like their being able to present many relations in a simultaneous way. Thus, it seems that the realist should provide an account of how diagrammatic reasoning is consistent with (R1) – in particular, how we come to know the ideal mathematical objects with the use of spatial representations, explaining at the same time how their use relates to empiricist methodology (which will bear on (R4))[17]. The shape of such an account will depend on the view on the nature of mathematical objects, but also their relation to mathematical language. Summing up, I hope to have shown that in giving any realist account of the nature of mathematical knowledge and mathematical objects, an explanation of how diagrammatic reasoning fits into the picture is also necessary.

## References

1. Brown, J.R.: Philosophy of Mathematics: An Introduction to the World of Proofs and Pictures. Routledge, London/New York (1999)
2. Coliva, A.: Human diagrammatic reasoning and seeing-as. Synthese **186**, 121–148 (2012)
3. Giaquinto, M.: Visual Thinking in Mathematics. Oxford University Press, Oxford/New York (2007)
4. Greaves, M.: The Philosophical Status of Diagrams. CSLI Publication, Stanford (2002)
5. Mancosu, P.: Philosophy of Mathematics & Mathematical Practice in the Seventeenth Century. Oxford University Press, New York/Oxford (1996)
6. Nikulin, D.: Imagination and Mathematics in Proclus. Anc. Philos. **28**, 153–172 (2008)
7. Plato: The Republic. Basic Books, New York (1968)
8. Plato: Letter VII. http://classics.mit.edu/Plato/seventh_letter.html
9. Sherry, D.: The role of diagrams in mathematical arguments. Found. Sci. **14**, 59–74 (2009)
10. Shimojima, A.: Semantic Properties of Diagrams and Their Cognitive Potentials. CSLI Publications, Stanford (2015)
11. Stenning, K.: Seeing reason: image and language in learning to think. Oxford University Press, Oxford (2002)

---

[17] However, it seems that the character of diagrammatic reasoning has little relevance to theses (R2) and (R3), which relate to the objective and necessary nature of mathematical knowledge.

# Diagram Design, Principles and Classification

# Picturing Science: Design Patterns in Graphical Abstracts

Jessica Hullman[1](✉) and Benjamin Bach[2]

[1] University of Washington, Seattle, WA 98115, USA
jhullman@uw.edu
[2] University of Edinburgh, Edinburgh EH8 9AB, UK
bbach@inf.ed.ac.uk
http://faculty.washington.edu/jhullman
http://benjbach.me

**Abstract.** A graphical abstract (GA) provides a concise visual summary of a scientific contribution. GAs are increasingly required by journals to help make scientific publications more accessible to readers. We characterize the design space of GAs through a qualitative analysis of 54 GAs from a range of disciplines, and descriptions of GA design principles from scientific publishers. We present a set of design dimensions, visual structures, and design templates that describe how GAs communicate via pictorial and symbolic elements. By reflecting on how GAs employ visual metaphors, representational genres, and text relative to prior characterizations of how diagrams communicate, our work sheds light on how and why GAs may be distinct. We outline steps for future work at the intersection of HCI, AI, and scientific communication aimed at the creation of GAs.

**Keywords:** Graphical abstract · Diagram · Information visualization

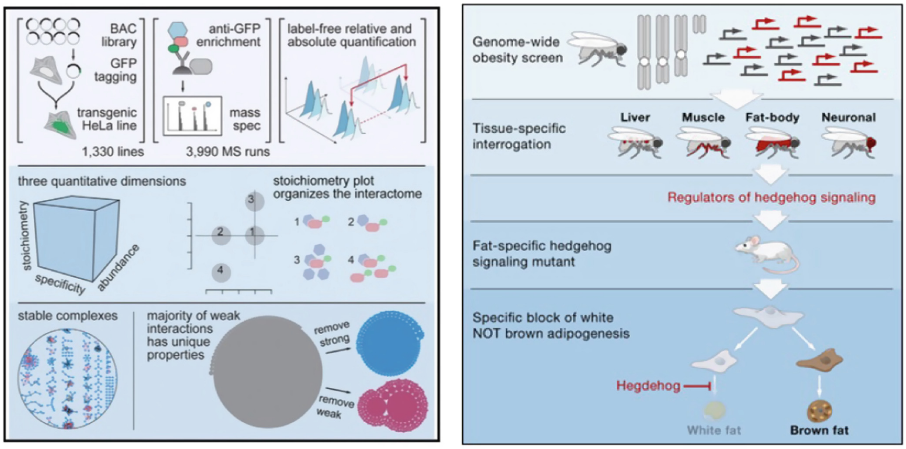
## 1 Introduction

The overwhelming scale of scientific publishing—partly accelerated though digital publishing [15]—increases the number of articles that must be consulted in the research process. In an effort to make it easier for readers to grasp the gist of publications, multiple scientific publishers have mandated that authors prepare a graphical representation of their primary findings. This form of Graphical Abstract (GA) represents a "single, concise, pictorial and visual summary of the main findings of the article" [5] (Fig. 1). By leveraging the efficiency of visual communication for portraying the essence of complex information, the assumption is that GAs will make scientific publications more accessible and understandable for in- and out-of-domain researchers as well as "lay" audiences like students, journalists, or members of the public.

Graphical abstracts are becoming more common, either as a requirement or suggestion, among journals spanning scientific domains. Several scientific journals have been publishing GAs for nearly a decade [24]. However, relatively little is known about what visual and textual mechanisms GAs employ, and how,

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 183–200, 2018.
https://doi.org/10.1007/978-3-319-91376-6_19

to express scientific research. While prior studies of diagrammatic communication can inform understanding of GAs (e.g., [8–10,33,34,36]), GAs are unique based on their focus on communicating scientific contributions specific to single publications. GAs can be thought of as a specific form of overview figure—a summative diagram used to aid readers in deciphering research contributions and methods across many scientific and empirical disciplines. Consequently, a deeper understanding of GAs is beneficial outside the specific cases in which they are required. The fact that scientists—who are not typically trained in graphic communication—are responsible for creating GAs means that a better understanding of the design space of these diagrams could lead to important contributions to design support. More specific design guidelines can be proposed than the high level suggestions currently provided by publishers [5,25,41] (e.g., "emphasize the new finding from the paper" [25]). Additionally, better authoring tools for GA design could help address the huge body of existing publications that lack GAs and better support the needs of those creating them. Techniques from automated and mixed-initiative graphic design (e.g., [22,23]) could be adapted to support the unique needs of scientists as designers.



(a) (b)

**Fig. 1.** Examples of layouts in GAs. (a) GA using a zigzag layout with illustrations and data visualizations to depict a research method for studying protein interactions. (b) GA combining elements of a unidirectional and a forking layout to illustrate findings about the fly genome. (Color figure online)

We are interested in identifying a repertoire of common structures and patterns in GAs. We believe that many GA authors arrive at patterns (e.g. choice of layout) without necessarily understanding how a given pattern relates to their communication goals. By surfacing common patterns, our work can contribute to helping authors of GAs to make more informed decisions in the future. We contribute the findings from a qualitative analysis of a sample of 54 GAs drawn from a range of scientific disciplines. Our primary contribution is an analysis of the pictorial, symbolic, and textual elements used in GAs. Our analysis characterizes design patterns associated with common GA design choices including the use of spatial layout, the type of picture or visual representation, how time is visualized, and how text is incorporated. We identify underlying design dimensions, including the linearity of spatial layouts, the degree of iconicity in picture types, and the degree to which text stands for versus narrates content. We describe how our results confirm, as well as problematize, prior conceptions of how scientific diagrams convey meaning. We reflect on how the unique status of scientists as graphic designers may impact the effectiveness of GAs, and present a set of design recommendations based on our study results. We conclude by motivating a research and development agenda for promoting the more effective design and use of GAs. We contrast the current state of GAs with a set of possible (imagined) functions of GAs as a form of scientific communication. We describe how research in Human-Computer-Interaction (HCI), Artificial Intelligence (AI), and science communication can be applied to facilitate the creation of effective GAs across disciplines.

## 2 Related Work

As a first step toward understanding the design of GAs, Yoon and Chung [42] recently examined the frequency and forms of GA use in the social sciences. Their coding process ("tagging") differentiated the scientific discipline, basic form of the GA (e.g., tables, charts (data visualizations), and diagrams), the content of the GA (e.g., the background, method, or results of the research), and whether the GA content was newly created or taken from an existing visualization in the publication. Their analysis surfaces patterns in overall forms of representations used and their relation to the content of the GA. Namely, schematic diagrams were most commonly used for presenting methods, background, or overviews of research, while data-driven charts were slightly more common than diagrams for depicting research results. Based on the prevalence of diagrams, our work takes a deeper look at the rich pictorial and symbolic space of GAs, which we find often transcends simple notions of representational genre (e.g., a large space of possible image schemas can be utilized within a diagrammatic GA). We also analyze a sample of GAs that includes a larger range of scientific disciplines.

The study of diagrammatic communication can inform our reading of how basic visual structures convey meaning in GAs. At a high level, diagrams schematize thought, using place and forms in space to convey both concrete and abstract meanings [34]. Certain "privileged" symbols (arrows, lines, boxes, crosses, and circles) and dimensions of the page (horizontal, vertical) are commonly used to convey meaning in diagrammatic contexts [33,36]. For example, arrows have been found to strongly imply the functional (as opposed to the structural) organization of mechanical systems (e.g., the temporal, dynamic, and causal aspects) [10]. Graphic space itself is described as naturally conveying relations being elements, including nominal, ordinal, and interval and ratio relations [32].

Temporal information has been found to be more commonly mapped to the horizontal dimension [35]. The prevalence of temporal processes in scientific methods makes techniques and interpretive models of diagrammatic depictions of change such as Arnheim's notion of implied motion [1] relevant.

More generally, a number of visual metaphors, or image schemas, have been identified by cognitive psychologists and others [2,11,14]. Image schemas, such as "more is up, less is down," center-periphery, or containment are apparent in language as well as visual communication [14]. Image schemas are thought to metaphorically structure our thinking pre-conceptually [14]. As a result, space is "not neutral", even to children [31]. Understanding the "logic" of these schema is critical to effectively using them in visual compositions like GAs to convey complex concepts. We are interested in more specific conventions in the use of spatial layout and other schematizing elements used to convey the contributions of single publications, as a point of comparison to prior studies of graphical communication focused on textbooks or other educational diagrams [32,38].

Another goal of our work is to identify how research on automated and mixed-initiative (i.e., human in the loop) design at the intersection of AI and HCI could be applied to support the design of GAs. Prior work in automated construction of diagrams visualizing scientific research has focused on producing a high level representation ideas in papers [27]. Approaches to representing individual research documents like PDFs include thumbnails of extracted images [4], or summary graphics that incorporate key terms and important images extracted from the paper [28]. However, these approaches rely on combining existing imagery from the publication, and cannot create a new, synthesizing representation.

## 3 GA Sample and Coding Process

As an initial step toward characterizing how GAs communicate scientific findings, we gathered a convenience sample of 54 GAs. Our goal was to build a sample that included some diversity in the visual structures employed, so as to serve our goal of demarcating a design space. We found that writing and guidelines about GAs often referenced examples that captured such diversity. We therefore seeded our sample with examples cited in prior writings about GAs, including examples included in design guidelines for GAs from scientific publishers [5,25] (20 GAs), examples contained in editorials about GAs [26] (1 GAs), and an example GA for a prior article about GAs (1, from [42]). As a secondary concern, we wanted to include GAs from varying disciplines. We added examples from papers retrieved through Google and Google Scholar searches on "graphical abstract" (14 GAs) and by browsing journal archives that require GAs (18 GAs), continuing to select examples that were diverse in visual structures and discipline.

We employed an open coding approach [17] informed by past analysis of graphical abstracts [42], visual semiotics [18], and visual structures common to diagrams and metaphor [2,10,11,14,33,34,36]. Each author first independently analyzed a subset of the sample (6 GAs), making notes of the visual structures, text, and other elements used in the GA, and what higher level dimensions best described distinctions between GAs. For context, we read the research publication associated with each GA. Through discussion, we arrived at an initial set of codes related to the spatial composition or layout of GAs and their use of symbols, pictorial elements, and text. Through further parallel coding of a new subset (10 GAs) and subsequent discussion, we refined these codes. We merged categories that seemed redundant and creating new categories to capture emerging dimensions of the design space. We repeated this process three more times until our scheme stabilized around four high level design aspects, each of which was associated with 4–9 individual codes. We then divided the GAs and each author independently coded half in a final coding. All codes were discussed and agreed upon.

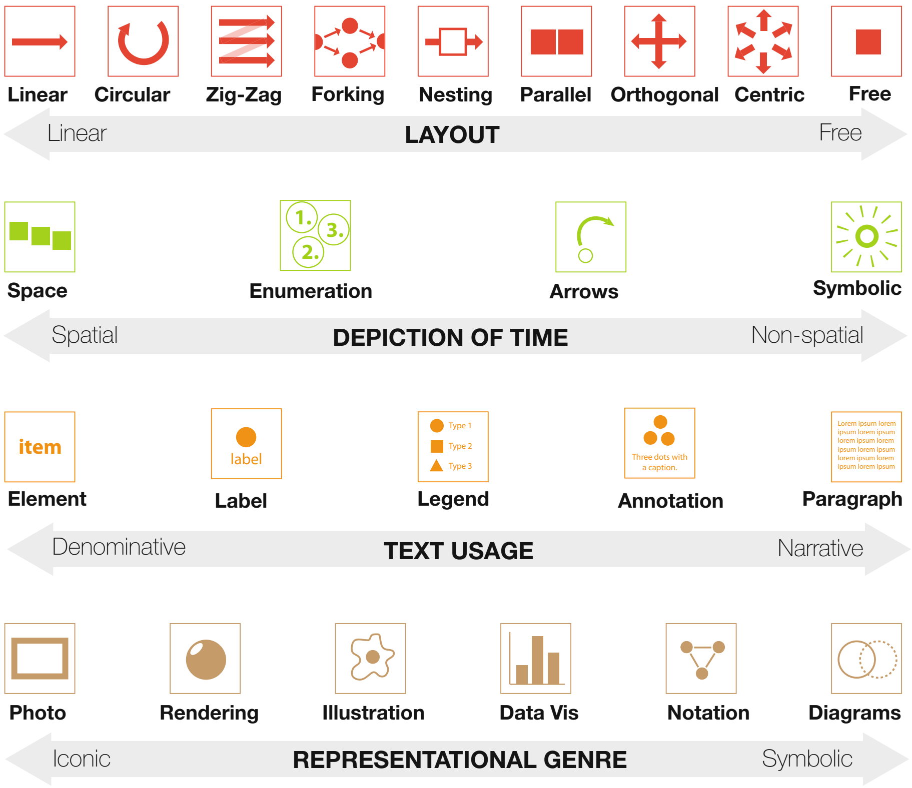


**Fig. 2.** Design patterns organized into four aspects: layout, depiction of time, text usage, and representational genre.

## 4 The GA Design Space

At a high level, our coding scheme acknowledges the way in which GAs employ graphical entities and text to communicate features of the research content at various levels of abstraction. Prior work describes how diagrams can convey both functional and structural information [10]. How the space of a diagram is used (structure), and the reader's eye is guided in a sequential way (function), are often strongly influenced by the content being depicted. However, authors of diagrams also face a number of more arbitrary choices about how abstractly they wish to communicate their content. For example, text is highly abstract in how it refers to meanings, whereas photographs are highly concrete.

Our coding scheme operationalizes these aspects of both content and communication style by defining four *design aspects* to capture the diversity of visual mechanisms that GAs employed in our sample (Fig. 2):

- **Layout** describes the organization of graphical elements in the 2D space of the GA. Layouts vary in the degree to which they imply a reading order (*linear* to *free*). Layout result from symbols (e.g., arrows), spatial mappings (e.g., organizing visual elements to extend from the center of the composition outward), or other implied relationships between elements (e.g., nesting one picture inside another).
- **Depiction of time** describes how the GA conveys a temporal process, a common function we observed in many GAs. With layout, depiction of time tends to be influenced by the scientific content that a GA is intended to communicate.
- **Text usage** differentiates ways of incorporating text in GAs, such as labels, paragraphs, or annotations (*denominative* to *narrative*). Most GAs we observed combined text with visual elements. The choice of whether to use text versus visual means of communication is also relatively independent from the scientific content being expressed than choices related to layout or time depiction.
- **Representational genre** describes types of representations that comprise a GA. These can vary in their degree of abstraction and style, and include photographs and screenshots, illustrations, scientific visualizations, abstract data visualizations, and schematic diagrams (*iconic* to *symbolic*). Representational genres can be correlated with the content of the research displayed in a GA in some cases (e.g., microscopy-based studies will more often present images in a GA). However, in many cases representational genre is relatively independent from the content being displayed.

Codes were not mutually exclusive: a GA could demonstrate multiple codes (design patterns) associated with the same aspect (e.g., layout, time). The frequency of each design pattern in our sample is indicated in the respective section.

## 5 Design Patterns

We describe the results of our analysis in terms of sets of *design patterns* we observed: standard solutions employed to solve a specific design problem. We identified distinct sets of design patterns associated with the design aspects of layout, depiction of time, text usage, and representational genre. We include example GAs and provide references to additional GAs in our sample which we include as supplemental material (*SM*[1], [2]).

## 5.1 Layout

Layout design patterns describe ways in which an author can use the graphic space of a GA to represent relations between pictures and concepts. As the fundamental means of organizing the space of a GA, layout is critical: prior work describes how cognitively, space is "not neutral", even to children [36]. How scientists choose to use layout sets up the implied "logic" that a GA conveys about a piece of research, whether intentionally or accidentally. The layouts we observed can be organized along a continuum, from *linear* layouts that convey an explicit reading order, to *non-linear* layouts that can be read in various orders (Fig. 2 top). Whether a GA layout is more linear or parallel is a function of how the layout uses common symbols (e.g., arrows or boxes) as well as spatial schemas (e.g., center-periphery, linear) to relate the components of the GA.

**Linear** (**19 GAs, 35%**): At the linear end of the spectrum, we observed simple linear layouts that used arrows or other visual cues to designate a clear reading order.

**Zig-zag** (**3 GAs, 5.5%**): Some layouts imply reading order through the use of both horizontal and vertical space. A zig-zag layout (e.g., Fig. 1(a)) consists of multiple rows, each of which implies the same horizontal reading order. Typically these layouts are designed to be read left-to-right and top-to-bottom [39].

**Forking** (**10 GAs, 18.5%**): Other layouts include both cues to a linear reading order and cues that break the implied linearity. For example, we observed multiple GAs using forking layouts, where entities or nodes are connected with paths (typically arrows) (e.g., Fig. 1(b), bottom half). In a forking layout, the process being represented branches in multiple directions at least once, such that it is not strictly linear.

**Nesting** (**16 GAs, 29.6%**): Nesting is a common layout strategy that can also conflict with cues to linear reading order. Nesting in GAs is analogous to footnoting in text, where a nested frame within the larger GA composition contributes additional contextual information that would not otherwise have appeared. For example, nesting can be used to emphasize an intermediate step in a process by relegating it to a separate frame than the rest of the depiction (e.g., *SM#51*).

A common use of nesting is to incorporate representations that vary the scale or level of detail. The difference in scale is depicted by a boundary around the nested frame(s) using either color, borders, or other visual differentiation (e.g., dotted circles in Fig. 3(a)). When used to vary scale or level of detail, nesting

---

[1] http://faculty.washington.edu/jhullman/GA_Sample_Images_Source_Info.pdf.

[2] http://faculty.washington.edu/jhullman/GA_Table.pdf.

can problematize an otherwise clear reading order, by adding ambiguity about which level of detail should be examined first.

**Parallel** (**20 GAs, 37**%): Layouts that used nesting could also fall near the parallel end of the spectrum, such as when several frames were juxtaposed, each with a nested component (e.g., Fig. 3(a)). Parallel layouts consist of side-by-side or vertical juxtapositions of multiple alternatives (which might be structures, processes, outcomes, etc.) for comparison, or multiple different representations juxtaposed without a clear indication of reading order (e.g., Fig. 3(b)).

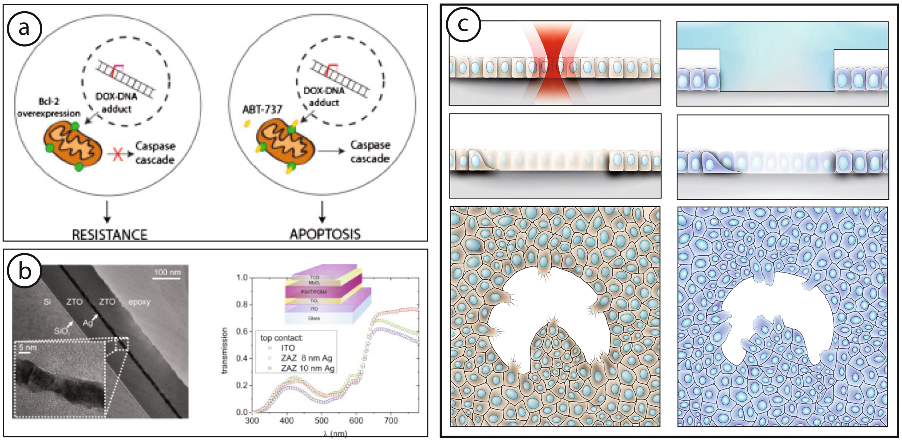


**Fig. 3.** GAs showing *parallel* (a and b), *orthogonal* (c), and *nested* (a) layouts. (Color figure online)

**Orthogonal** (**9 GAs, 16.6**%): Similar to parallel layouts, orthogonal layouts do not always indicate one clear reading order, but map different information to both the horizontal and vertical dimensions of the GA (e.g., Fig. 3(c)). Orthogonal layouts impose a grid on the space of the GA, which, in the absence of indicators of reading order, is analogous to the "anti-narrative" effect of grids in modern art [13]. Visual cues as to the hierarchy of the four quadrants may counteract the ambiguity, and conventional reading orders based on culture may compel readers to adopt certain orders. However, the lack of explicit order indicators in the GA design nonetheless subtly implies the equivalence and non-temporal relationships among the information being depicted.

**Centric** (**4 GAs, 7.4**%): Centric layouts are have ambiguous reading orders. A centric layout divides the space of the GA into a center and a periphery, typically mapping elements to both types of position such that it is unclear which direction should be examined after the center.

SINGLE (**8 GAs, 14.4%**): Finally, some GAs used a single layout, consisting of a picture that was not clearly differentiable into sub-pictures. We observed single diagrams (e.g., *SM#4*), single data visualizations (e.g., *SM#37*), and single photographs (e.g., *SM#41*).

**Vertical Versus Horizontal Dominance.** Prior work describes how graphical displays in the sciences tend to be remarkably constrained in their use of space, relying most heavily on vertical space unless a neutral dimension like time is shown. For instance, a prior analysis of scientific diagrams in textbooks provided quantitative evidence of the dominance of vertical arrays over horizontal. Only 2 out of 48 charts (4%) found in biology, geology, and linguistics textbooks in the Stanford Undergraduate Library used the horizontal dimension as the primary organizing direction [32]. We compared the frequency with which GAs in our sample utilized the vertical, horizontal, or both dimensions for comparison. Only three of the 17 GAs in our sample that relied on a single dimension used the vertical dimension. Two of these GAs used vertical layout to depict a non-evaluative dimension, including to display several alternative models and several graphical representations of results. Of the much larger proportion of GAs in our sample that utilized both vertical and horizontal layout (28/54 or 52%), several used visual cues to prioritize the vertical dimension while using the horizontal dimension to show alternative views as might be predicted by the prior work. However, three of these 28 GAs used vertical layout to depict time, which prior work suggests is typically mapped to horizontal layout [32]. The 14 GAs that used only horizontal layout were, on the other hand, more likely than not to use space for a neutral dimension like time or to show alternatives.

## 5.2 Depiction of Time

A majority of GAs (83%) depicted processes, including both natural processes like cellular division or engineered processes like technical pipelines. Processes are by definition temporal, requiring strategies for representing time in the 2D graphical plane. Of all 54 GAs in our sample, only 9 (17%) did not represent any temporal information. We observed GAs employing several specific strategies to depict temporal information spanning symbolic and spatial approaches. The most prevalent depictions of time used arrows (72%); however, arrows necessarily involve a spatial mapping as well. Spatial mappings (without additional schematics to convey time) were also prevalent, with 54% of GAs using space.

SPATIAL (**29 GAs, 53.7%**): To map time to space is to essentially "unfold" a temporal process onto the 2D space of the GA. The steps in a process or temporal snapshots of a system can be represented in pictures like visualizations, schemas, or photographs, and laid out with a specific reading ordering. The implicit reading order in western culture is left-right and top-down, though the presence of other visual cues such as arrows can change the implied order. These types of spatial mappings can use any of the linear layouts described in Sect. 5.1.

Prior work suggests that the horizontal dimension is more frequently used to depict time in diagrams [35], as an example of a more general convention of using the horizontal to depict neutral, as opposed to evaluative, dimensions [32]. Among those 29 GAs that used space to convey a temporal process, 13 GAs (45%) used the horizontal dimension, and 4 GAs (14%) used the vertical dimension. The remaining 12 GAs used both dimensions.

**Enumeration** (**4 GAs, 7.4**%): Enumerations such as roman numerals (i, ii, iii), letters (a, b, c) or others can reinforce the reading order in spatial mappings where it is otherwise ambiguous.

**Arrows** (**39 GAs, 72.2**%): More common than enumerations are arrows, a widely used symbol to indicate the intended direction in which visual elements should be examined. We observed GAs using arrows in two ways: 1) to indicate sequence between pictures in a spatial mapping (e.g., *SM#7*), and 2) to indicate direction of movement or action in a depiction of a dynamic process (*SM#3*).

**Symbolic** (**5 GAs, 9.3**%): Other forms of symbolic representations for depicting time are relatively rare in GAs. For example, change can be depicted through blur atop a changing object (e.g., *SM#29*) .

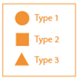
### 5.3 Text Usage

Many GAs combined visual representations with text. Research indicates that combining visual and text modalities promotes better understanding of complex phenomena, presumably because of the benefits of the cognitive work required to integrate information across modalities [19]. In particular, prior work has shown that to clearly convey a process often necessitates both text and diagrams [8].

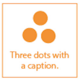
We observed a range of uses of text in GAs. Text served multiple functions across GAs in our sample, ranging from concise use of text to *denote* objects or processes that were not otherwise represented, to longer descriptions used to *narrate* or reason about a depicted phenomena (similar to those studied in prior work on diagrams). Text usage also varied in how "anchored" the text was to the visual element. For example, labels were clearly anchored to their referents, while the intended referent of a commentary was generally ambiguous.

**Index** (**8 GAs, 14.8**%): Some GAs substituted text for visual representations of an entity. In these examples, text is used as an index, either for an organism or substance (e.g., Fig. 1(b)) or process (e.g., Fig. 1(a)).

**Label** (**43 GAs, 79.6**%: The most common use of text in the GAs in our sample is to label a visually-represented object, process, or state in a process (e.g., Fig. 3(a)). When used as labels, the primary function of the text is to name. Labels can describe simple atomic pictures (e.g., *SM#7*, *SM#20*) or more complex composites comprised of multiple pictures (e.g., Fig. 1(a) 'stable complexes'). Some GAs labels color coded labels so that they perform the additional function of associating objects, process states or labels with one another (e.g., Fig. 4).



**Legends** (**11 GAs, 20.4%**): Legends provide global explanations, explaining symbols in visual notation (e.g., *SM#9*) or visual encodings in data visualizations (e.g., Fig. 3b).



**Caption** (**37 GAs, 68.5%**): Some GAs used text to describe a set of visually-represented entities, as opposed to simply naming individual elements. These captions varied from providing concise descriptive information such as measurements (e.g., Fig. 1(a) '3,990 runs') to complete sentences describing processes or states (e.g., Fig. 1(a) text in center and bottom panels). Captions varied in how explicitly they were related to the representations they described. For example, some GAs used multiple parallel frames, each had a similarly-styled caption in the same location from the frame (e.g., *SM#23*, *SM#50*). In other cases, captions are more implicitly associated with sets of representations through proximity (e.g., Fig. 1(a) center left).

**Commentary** (**3 GAs, 5.6%**): While captions were used to describe distinguishable subsets of a GA, text could also be used to describe without clearly referencing parts of the GA. Instead, text as commentary added explanation or context for the GA without any apparent anchor (e.g., *SM#23* center).

Only 3 of the GAs in our sample used no text.

## 5.4 Representational Genre

Most GAs combine two or more types of representations, each conveying a particular type of information. Examples include photographs, hand-drawn illustrations, computer generated visualizations, or schematic representations. Each representational genre may be associated with a plethora of subtypes, e.g., different types of data visualizations or different types of schemata. For our purposes of describing the design space of GAs, we are primarily interested in the broader distinctions between the genres used, including what types of information each conveys.

We differentiate representational genres according to their Iconicity, the degree to which they show real world elements or conceptual and abstract ideas. Structuralists differentiated between *icons* as precise representation of real objects, and *symbols* being abstract visual constructs (e.g. cross, circle) but referring to general ideas (e.g., Christianity, women's bathroom) [20]. A high iconicity representation focuses on a

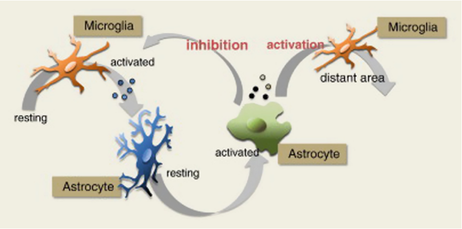


**Fig. 4.** A GA using a schema to show activation patterns in the nervous system. (Color figure online)

specific real world entity such as a specific cell (*SM#3*) or crystalline formation (*SM#19*). Representations with high iconicity depict real world objects that are potentially visible to the human eye and captured through cameras or microscopes. These representations can be used to imply the individuality or detail of the entity pictured. On the other side of the spectrum are pictures with a low iconicity: schemas depicting abstract or general constructs such as center periphery (e.g., *SM#22*) or flow diagrams (e.g., Fig. 4).

The representational genre—such as photo, scientific visualization, illustration, data visualization, symbolic notation, and schema—is in many cases influenced by the information to be shown. However, the level of iconicity can imply information and meaning beyond what is shown. For example, a visualization of an fMRI scan can refer to a single patient, implying an exemplary or abnormal case. A hand-draw illustration of the same "data" can imply a class of cases.

GAs in our sample employed a spectrum of representations that lay between photos and schemas in iconicity. We describe these in order of decreasing iconicity.

**Photographs (7 GAs, 13%)**: Photographs include photographs taken with an ordinary camera (e.g. or a study setup), or through a microscope (e.g., Fig. 3(b) left).

**Scientific visualizations (7 GAs, 13%)**: Scientific visualization is concerned with the representation, rendering, and exploration of intrinsically "spatial" three-dimensional data: anatomical body scans, particle flows, architecture, or machinery. The visualizations are intended to faithfully represent these objects to allow exploration and analysis of their structures; for example, brain tumors or functional MRI data. Scientific visualizations have become a standard method in many scientific domains. Their iconicity is high; however, as the data may be incomplete (sampling rate, sampling errors), it is possible that detail about the real world object may have been lost. In GAs, scientific visualizations have been used to show MRI (e.g., *SM#14*) and other anatomical data (e.g., *SM#15*, *SM#16*), as well as 3D protein structures (e.g., *SM#19*) and tectonic structures (e.g., *SM#35*).

**Illustrations (37 GAs, 68.5%)**: Illustrations are hand-drawings (using tools such as, e.g., Adobe Illustrator) showing objects in somewhat higher abstraction and less detail. Examples include animals (e.g., Fig. 1(b)), cells (e.g., Fig. 3(a)), and tools (e.g., Fig. 3(c), top). Illustrations are not meant to replicate specific real-world elements, but to represent a class of these objects, or the general idea: e.g., "hormone injected in mouse". Illustrations appearing in our GA sample were often used to show processes such as the interaction between biological entities (e.g., *SM#1*, *SM#3*) or to illustrate a specific research methodology (e.g., Fig. 1(b)).

**Data Visualizations (11 GAs, 20.4%)**: Data visualizations are representations of abstract data, i.e., data that is not associated with an inherent three-dimensional representation. Abstract data involves numeric values such as scientific measures and statistics, but can also refer to more complex data structures such as trees (e.g. taxonomies), networks, and temporal data. GAs in our sample used data visualizations in the context of measurement presentation (charts in Figs. 1(a); 3(b) right) and gene expression levels (*SM#24*), among others. Data visualizations can be snapshotted directly from a visualization program (e.g. python), or further abstracted by recreating using, e.g., Adobe Illustrator.

**Symbolic Notation (13 GAs, 24.1%)**: Symbolic notation refers to graphical codes that use mostly domain specific symbols and compositions. Symbolic notation is used, e.g., to depict chemical molecules (e.g., *SM#32*) or convey information about genes (e..g, symbols on the bottom line of the green charts in *SM#24*). Symbolic notation is similar to data visualization but rather than showing a particular real-world instance (that the data is describing), these symbols often express non-existing concepts (e.g., *SM#17*) and even processes (e.g., *SM#24*).

**Schemas (29 GAs, 53.7%)**: Schemas are the most abstract representational genre used in GAs. Schemas employ various common symbols and spatial mappings to express ideas, concepts, and processes. Schematic elements can span an entire GA, effectively turning the entire GA into a schema (e.g., Fig. 4). More than other genres, schemas employ layout devices that help convey the logic behind the depiction (e.g., linear to denote a process, parallel to denote alternatives, etc.).

## 6 Discussion

### 6.1 Scope and Results

Our analysis is based on 54 GAs. While many journals require GAs, identifying a diverse sample of GAs is challenging. We aimed to retrieve GAs from a variety of domains our sample remains dominated by GAs from biology and chemical sciences. However, we believe the design patterns that we identified are not specific to any domain. Moreover, we suspect that our design patterns could also be applied to other genres of scientific presentations such as posters, infographics, or data comics [3]. As GAs become a more common requirement, future work should seek more reliable ways of collecting graphical abstracts across disciplines.

The patterns and dimensions that our analysis identified (layout, depiction of time, text usage, and representational genre) were arrived at through considerable discussion. These patterns and dimensions represent those that both authors determined best discriminated between differences in GAs in the sample. Additional dimensions could include color or visual styling, though we found these dimensions where less related to the content and the message of the graphical abstract.

While various publishers [5,25,41] and online guides [6,37] provide high level design advice to GA authors (e.g., “Use subtle colors” [6,25], “use captions” [41], “designate a clear reading order” [5,25]), it may be unrealistic to expect scientists who have little prior experience or training in graphic design to implement these guidelines. Our patterns provide a much more concrete starting point to assess what makes for an efficient and effective GA. Our framework can facilitate the process of developing guidelines; e.g., the effectiveness of a particular pattern can be assessed for different intentions (e.g., unidirectional layout for demonstrating sequential steps) or critiqued from a perspective that assumes more general goals like designating a clear reading order and clearly conveying the contribution of the research.

### 6.2 Ambiguity in the Design of GAs

Many prior studies have focused on professionally created diagrams such as those appearing in textbooks [30,32]. Relative to Tversky's [32] study of charts in textbooks, the design of GAs is more diverse in its use of spatial layout than the textbook diagrams, which were presumably created by professional artists.

Only 27 out of the 54 GAs in our sample used a vertical organization, and only three of those GAs used a vertical organization alone. Instead, many of the layouts used in GAs in our sample that made use of both vertical and horizontal space (28/54 or 52%) lacked visual features that would prioritize one of the dimensions to guide reading. Figure 5 depicts one such GA. Similarly, of the GAs that used spatial layout to depict time, while horizontal time was more common (by roughly 3 to 1 odds) only 4 GAs used vertical alone, while the others used both dimensions, lending ambiguity in reading order in some cases.

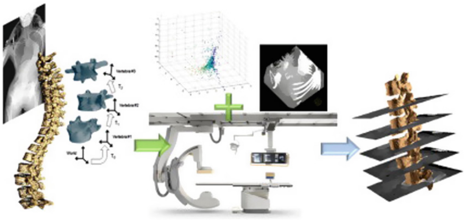


**Fig. 5.** A GA published with a paper on an inference method for articulated spine model [12], presented as an exemplar GA by Elsevier [5].

In addition to unclear reading orders, we observed various other violations across multiple GAs in our sample. **Unclear relationships between pictures** make it ambiguous whether multiple pictures refer to a natural, temporal sequence, or a methodology based on human intervention, or whether they represented equal alternatives. **Missing annotations** make it hard to explain and interpret visual elements in illustrations or pictures, including data visualizations. **Inconsistent visual styles** for text and graphical elements raise questions about whether the differences are intended to convey information or are arbitrary.

### 6.3 Toward Mixed-Initiative GA Design Tools

We suspect that many GA authors currently use general visual design tools, like Adobe Illustrator, Adobe Photoshop, or even simpler tools such as MS Powerpoint and Keynote. Authors in certain disciplines are likely to use domain specific programs to create diagrams (e.g., molecular structures). However, the

GAs that we observed that appeared to be created using specialized programs were more likely to lack clear labeling, reading order, and other visual cues to support reading, perhaps because the program used does not offer or emphasize such functions. To enable effective GA design among the many untrained graphic designers who are tasked with creating GAs may require a new breed of GA authoring tools that are customized to the communication intents of GAs.

Design templates that organize or style information according to different patterns or themes (e.g., [29,40]) are one possible solution for lowering the barrier for visual design of GAs. The design patterns we identify suggest a set of *design templates*, common configurations of pictorial and symbolic elements, that might be used to constrain the large space of possibilities available to authors. For instance, patterns that we observed which seem indicative of differences in the style of research include:

- **Process illustrations:** The majority of GAs depicted processes using (handcrafted) illustrations on a linear or forking layout. Labels were used to name elements while arrows connected between the individual stages and pictures.
- **Result representations:** Many GAs incorporated data visualizations to depict research results in the form of line charts (e.g., Fig. 3(c)), often in the context of a larger flow diagram or another spatial mapping used to depict a temporal process.
- **Parallel layouts:** Some GAs consisted of parallel pictures laid out horizontally on the GA plane (e.g., Fig. 3(c)). These GAs were frequently associated with survey and review type articles, as well as research that contributed a comparison between alternative forms of a process or structure. In these examples, both pictures are intended to be of equal importance, with no indication of sequence between them.

We envision a system that presents a designer with an initial choice of templates like those above. Such a system could integrate examples (e.g., GAs plus text abstracts for context) to better allow an author to identify the right pattern for the type of contributions their work makes. Exposure to examples in the design process can help designers realize important structures and transfer these to other situations [7], and has been shown to improve design quality [16]. Given techniques for extracting and sufficiently annotating GAs with contextual information, authors could search a GA example library directly or receive automatic suggestions of relevant designs in the design process from GAs expressing similar research contributions or visual elements.

*Mixed initiative tools* are those in which a human creator is given access to system recommendations to help his or her work. Mixed initiative tools for visual design have included authoring tools for single page graphic designs like posters and flyers (e.g., [23]). As complex visual compositions that contain schematic elements, photographs, visualizations, and text, GA designers could likely benefit from system suggestions regarding different design aspects.

For example, a template and example-oriented approach can be combined with automated suggestions to improve features of a GA as an author creates it. Interactive layout suggestions, including changes in the position, scale, and alignment of elements, have been shown to help novices produce better quality designs, as rated by other novices [21]. Such suggestions can include refinements aimed at improving an author's current design, or larger proposed changes to the style of a design, including the layout [21]. A "design validator" tool could allow GA authors who are not confident in their graphic design skills to get suggestions and feedback on aspects of a design like font choices [22] or color choices.

Automatic tools that generate GAs by extracting content and structure from a scientific article are another possible direction. For example, the DocumentCards [28] creates a visual summary of a paper. However, DocumentCards rely on a simple "formula" designed to fit the structure of the average research paper, and cannot take into account differences between the research content or purposes of papers. The difficulty of extracting an expert's notion of which contributions of a work are critical or how they differ from other works suggest that a hybrid approach combining judgments from a human scientist with automatic features may be most effective for facilitating GA design.

## 7 Conclusion

Graphical Abstracts (GAs) are increasingly required by publishers to make scientific findings more accessible across and within disciplines. We contributed the first analysis of the pictorial and symbolic design space of GAs. By applying visual communication knowledge and qualitative coding, our analysis aims to pave the way for future empirical work and authoring systems focused on GA design. We identified four design aspects–*layout*, *depiction of time*, *text usage*, and *representational genre*–differentiating a range of design patterns associated with each. We describe how the design choices made in GAs point to common design templates and the design advice they imply. We outline directions for future study and development to facilitate GA design.

## References

1. Arnheim, R.: Art and Visual Perception: A Psychology of the Creative Eye. University of California Press, Berkeley (1954)
2. Arnheim, R.: Visual Thinking. University of California Press, Berkeley (1969)
3. Bach, B., Wang, Z., Farinella, M., Murray-Rust, D., Henry Riche, N.: Design patterns for data comics. In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI) (2018, to appear)
4. Bauer, D., Fastrez, P., Hollan, J.: Spatial tools for managing personal information collections. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences, p. 104b. IEEE (2005)
5. Elsevier: Graphical abstracts (2016). https://www.elsevier.com/authors/journal-authors/graphical-abstract
6. American journal Experts: How to make a good graphical abstract (2016). http://bitesizebio.com/31125/how-to-make-a-good-graphical-abstract

7. Gentner, D., Loewenstein, J., Thompson, L.: Learning and transfer: a general role for analogical encoding. J. Educ. Psychol. **95**(2), 393 (2003)
8. Hegarty, M., Just, M.A.: Constructing mental models of machines from text and diagrams. J. Mem. Lang. **32**, 717–742 (1993)
9. Heiser, J., Tversky, B.: Diagrams and descriptions in acquiring complex systems. In: Proceedings of the Meetings of the Cognitive Science Society (2002)
10. Heiser, J., Tversky, B.: Arrows in comprehending and producing mechanical diagrams. Cognit. Sci. **30**(3), 581–592 (2006)
11. Johnson, M.: The Body in the Mind: The Bodily Basis oOf Meaning, Imagination, and Reason. University of Chicago Press, Chicago (2013)
12. Kadoury, S., Labelle, H., Paragios, N.: Automatic inference of articulated spine models in ct images using high-order Markov random fields. Med. Image Anal. **15**(4), 426–437 (2011)
13. Krauss, R.E.: The Originality of the Avant-Garde and Other Modernist Myths. MIT Press, Cambridge (1986)
14. Lakoff, G.: Women, Fire, and Dangerous Things: What Categories Reveal About the Mind. Cambridge University Press, Cambridge (1990)
15. Larsen, P.O., Von Ins, M.: The rate of growth in scientific publication and the decline in coverage provided by science citation index. Scientometrics **84**(3), 575–603 (2010)
16. Lee, B., Srivastava, S., Kumar, R., Brafman, R., Klemmer, S.R.: Designing with interactive example galleries. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2257–2266. ACM (2010)
17. Lofland, J., Lofland, L.H.: Analyzing Social Settings. Wadsworth Publishing Company Belmont, Belmont (2006)
18. Margolis, E., Pauwels, L.: The Sage Handbook of Visual Research Methods. Sage, Thousand Oaks (2011)
19. Mayer, R.E.: Multimedia learning. Psychol. Learn. Motiv. **41**, 85–139 (2002)
20. Moriarty, S.: Visual semiotics theory. In: Handbook of Visual Communication: Theory, Methods, and Media, vol. 8, pp. 227–241 (2005)
21. O'Donovan, P., Agarwala, A., Hertzmann, A.: DesignScape: Design with interactive layout suggestions. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 1221–1224. ACM (2015)
22. O'Donovan, P., Lībeks, J., Agarwala, A., Hertzmann, A.: Exploratory font selection using crowdsourced attributes. ACM Trans. Graph. (TOG) **33**(4), 92 (2014)
23. O'Donovan, P., Agarwala, A., Hertzmann, A.: Learning layouts for single-pagegraphic designs. IEEE Trans. Vis. Comput. Graph. **20**(8), 1200–1213 (2014)
24. Pferschy-Wenzig, E.M., Pferschy, U., Wang, D., Mocan, A., Atanasov, A.G.: Does a graphical abstract bring more visibility to your paper? (2016)
25. Cell Press: Cell press graphical abstract guidelines (2016). http://www.cell.com/pb/assets/raw/shared/figureguidelines/GA_guide.pdf
26. Romans, B.: Are graphical abstracts a good idea? (2011). https://www.wired.com/2011/02/are-graphical-abstracts-a-good-idea
27. Shahaf, D., Guestrin, C., Horvitz, E.: Metro maps of science. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1122–1130. ACM (2012)
28. Strobelt, H., Oelke, D., Rohrdantz, C., Stoffel, A., Keim, D.A., Deussen, O.: Document cards: a top trumps visualization for documents. IEEE Trans. Vis. Comput. Graph. **15**(6), 1145–1152 (2009)
29. Adobe Systems: Adobe Spark Post (2017). https://spark.adobe.com/about/post

30. Tufte, E.R.: Envisioning information. Optom. Vis. Sci. **68**(4), 322–324 (1991)
31. Tversky, B.: Some ways that maps and diagrams communicate. In: Freksa, C., Habel, C., Brauer, W., Wender, K.F. (eds.) Spatial Cognition II. LNCS (LNAI), vol. 1849, pp. 72–79. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-45460-8_6
32. Tversky, B.: Spatial schemas in depictions. In: Spatial Schemas and Abstract Thought, pp. 79–111 (2001)
33. Tversky, B.: Visuospatial reasoning. In: The Cambridge Handbook of Thinking and Reasoning, pp. 209–240 (2005)
34. Tversky, B.: Visualizing thought. Topics Cognit. Sci. **3**(3), 499–535 (2011)
35. Tversky, B., Kugelmass, S., Winter, A.: Cross-cultural and developmental trends in graphic productions. Cogn. Psychol. **23**(4), 515–557 (1991)
36. Tversky, B., Zacks, J., Lee, P., Heiser, J.: Lines, blobs, crosses and arrows: diagrammatic communication with schematic figures. In: Anderson, M., Cheng, P., Haarslev, V. (eds.) Diagrams 2000. LNCS (LNAI), vol. 1889, pp. 221–230. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-44590-0_21
37. wikiHow Community: How to make a graphical abstract for scientific publication (2016). http://www.wikihow.com/Make-a-Graphical-Abstract-for-Scientific-Publication
38. Winn, B.: Charts, graphs, and diagrams in educational materials. Psychol. Illus. **1**, 152–198 (1987)
39. Winn, W.D.: A theoretical framework for research on learning from graphics. Int. J. Educ. Res. **14**(6), 553–564 (1990)
40. WordPress.org: Word press themes directory (2017). https://wordpress.org/themes/
41. IEEE Xplore: Graphical abstract description and specifications (2015). https://www.ieee.org/publications_standards/publications/graphical_abstract.pdf
42. Yoon, J., Chung, E.: An investigation on graphical abstracts use in scholarly articles. Int. J. Inf. Manage. **37**(1), 1371–1379 (2017)

# A Framework for Analyzing and Designing Diagrams and Graphics

Yuri Engelhardt[1](✉) and Clive Richards[2](✉)

[1] University of Twente, Enschede, The Netherlands
yuri.engelhardt@gmail.com
[2] Birmingham City University, Birmingham, UK
clive.j.richards@me.com

**Abstract.** A systematic method is presented that describes comprehensively the very broad design space of visualizations and the interrelationships between their constituents. The framework offered here includes three representational modes and fifteen visual encodings principles as the proposed universal building blocks of all types of diagrams and information graphics. The framework provides: 1. A vocabulary and a method for thoroughly analyzing the full spectrum of visual representations of information. 2. A mechanism for exploring previously unexploited combinations of visual encoding principles for representing information. 3. A potential tool for creating alternative representations for any given visualization or data set.

**Keywords:** Visual encoding · Types of visual representations
Design space · Building blocks · Graphic language · Graphic relationships
Nature of diagrams

## 1 Introduction

A number of authors have offered frameworks for the analysis and design of diagrams and information graphics, notably Bertin (1967, 1977); Twyman (1979); Johnson (1987); Lakoff (1987); Tversky (1995); Card and Mackinlay (1997) and Ware (2008). Each has contributed to our understanding of how information can be represented graphically. The present authors have reviewed these and other previous contributions, and identified gaps in what they cover. Previous frameworks:

1. Cover **only some aspects** of visual representation (such as depiction, visual encoding, mode of correspondence, etc., see Table 1).
2. Cannot be applied to all **types** of visualizations (e.g. clock faces, technical drawings, family trees, heatmaps, comic strips, 3D data sculptures, Isotype charts, etc.).
3. Cannot be used to show how every type of visualization can be constructed from the universal building blocks of diagrams and information graphics.

The framework we present here fills these gaps, offering a systematic way to describe comprehensively the very broad design space of visualizations and the interrelationships between their constituents (see also Richards and Engelhardt forthcoming). In the terminology of 'A taxonomy for diagram research' (Blackwell and

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 201–209, 2018.
https://doi.org/10.1007/978-3-319-91376-6_20

**Table 1.** Key terms of our framework and corresponding distinctions by other authors (due to different conceptual approaches, some matchings are necessarily approximate)

| *Richards–Engelhardt framework* | *Bertin (1967)* | *Twyman (1979)* | *Johnson (1987), Lakoff (1987)* | *Tversky (1995)* | *Card & Mackinlay (1997)* | *Ware (2008)* |
|---|---|---|---|---|---|---|
| **Mode of visual encoding:** | | | **Image schemata:** | | **Graphical properties:** | **Graphical codes:** |
| picturing | | non-linear | | | | |
| encoding by configuration: | **Organization of the plane:** | **Method of configuration:** | | **Spatial devices:** | | |
| mapping | quantitative difference in position | non-linear | | | position in space | |
| positioning along an axis | | pure linear [1 dim], matrix [2] | | degree of spatial proximity | | |
| ordering by position | ordered variation in position | pure linear, linear interrupted | linear order | [spatial] order | | spatially ordered |
| grouping by proximity | | | near-far | separation by empty space | | proximity |
| grouping by alignment | | list [1 dim], matrix [2 dim] | | | | |
| grouping by boundary | zones | | container | delineation | enclosure | enclosed by contour/color |
| connecting | correspondence lines | linear branching, non-linear | link, path | arrows | connection | connected by contour |
| nesting | nesting | | | | | nested regions |
| encoding by visual appearance: | **Retinal variables:** | | | | | |
| sizing | size | | scale | relative size | size | size, height, thickness |
| proportioning | | | [scale with] part-whole | | | |
| repeating | | | [scale with] iteration | | | |
| ordering by gradient | value | | | | gray-level | |
| grouping by color | color | | | color | color | same color |
| grouping by shape | shape | | | | shape | same shape |
| **Mode of correspondence:** | | | | | | |
| literal configuration | | | | spatial relations convey spatial rel. | | |
| visual appearance | | | | depicting the represented thing | | |
| non-literal configuration | | | | spatial relations convey other rel. | | |
| visual appearance | | | | figures of depiction, convention | | |
| **Mode of depiction:** | | **Mode of symbolization:** | | | | |
| realistic/precise | | pictorial | | detailed depictions | | |
| schematic | | | | schematized depictions | | |
| non-depictive | | schematic, verbal/numerical | | conventionalized symbols | | |

Engelhardt 2002) the scope of the framework proposed here covers the 'signs that are the components of a diagram', the 'graphic structure' of a diagram and the 'meaning' represented in a diagram, but not 'context-related aspects' (e.g. diagram use).[1]

This framework has its origins in *Diagrammatics* (Richards 1984) and *The Language of Graphics* (Engelhardt 2002). In synthesizing our separate investigations and developing them further, we have broadened the scope of our analysis. In addition to the collections examined previously we have now also analyzed specimens from corpuses including datavizcatalogue.com (60 visualization types) and datavizproject.com (154 visualization types).

Starting from an information design perspective the framework was produced in two steps. In step one we thoroughly reviewed and analyzed the distinctions drawn in the existing literature and applied them to the analysis of example specimens of diagrams and graphics, identifying overlaps, gaps and inconsistencies. Step two consisted of developing what we believe to be a more comprehensive system for identifying distinctive features and principles, in an iterative cycle of challenging the evolving framework by applying it to a broad range of example specimens, while continuously refining it.

We distinguish diverse aspects of visual representations. It is the relationships and dependencies between these aspects that form the core of the framework. Our key terms for these aspects are shown in Table 1, along with corresponding distinctions made by a selection of previous authors. In this paper our key terms appear in bold type when first used, and in italics in most subsequent appearances.

## 2 Graphic Relationships Between Graphic Components

Various authors have pointed out that relationships, both displayed and represented, are central to the nature of diagrams and graphics. According to the logician Charles Sanders Peirce "*A diagram is a representamen which is predominantly an icon of relations and is aided to be so by conventions*" (Peirce 1903, 4.418). In the 11th edition of the *Encyclopaedia Britannica* the physicist, James Clerk Maxwell (1910), defined 'Diagram' as: "*... a figure drawn in such a manner that the geometrical relations between the parts of the figure illustrate relations between other objects*". The cartographer Jacques Bertin noted that "*The aim of graphics is to make relationships ... appear. ... The transcription of relationships does not utilize 'signs'; it utilizes only the relationship between signs*" (Bertin 1977/1981, 176–177). And the philosopher Bertrand Russell observed that "*a map, for instance, is superior to language, since ... a relation is represented by a relation*" (Russell 1923, 84–92).

In line with this emphasis on relationships, a visual representation of information can be regarded as expressing meaning by way of **graphic relationships** between **graphic components**. *Graphic components* may be shapes, lines, symbols, pictures or

---

[1] Our framework applies to any 'visual representations (of information)', 'diagrams and (information) graphics' or 'visualizations' – terms which are used interchangeably here.

words. They may be organized into *graphic relationships* by their respective spatial positions, by having the same color or different sizes, by being connected by lines, etc.

A *graphic component* can be involved in several different *graphic relationships* at the same time, i.e. simultaneously involving several visual encoding principles, which we introduce below.

Sometimes a graphic component of a visualization may be regarded as consisting of graphic components at a more detailed level (which also express meaning through the graphic relationships between them), thus allowing for nested visual structures.

## 3 Modes of Visual Encoding, Depiction, and Correspondence

The way components within visual representations display graphically their relationships we describe as the **mode of visual encoding** (involving depicting, scaling, ordering, grouping, or linking).

Each *visual encoding* can be characterized by its **mode of correspondence** (being literal or non-literal). The *mode of correspondence* refers to how the relationships displayed may correspond to the relationships that are represented.

The **mode of depiction** (realistic/precise to schematic) describes how depictions may range from being spatially or visually realistic and precise to being spatially or visually synoptic and edited.

These three representational modes, *visual encoding*, *correspondence*, and *depiction* are now introduced in more detail, with a particular emphasis on the *mode of visual encoding*, the key feature of the framework presented here.

## 4 Mode of Visual Encoding

*Graphic relationships* in diagrams and graphics can be characterized by their *mode of visual encoding*. Five broad **types** of *visual encoding* can be identified (depicting, scaling, ordering, grouping, linking), which can be further broken down into fifteen **principles** of *visual encoding*. Principles of visual encoding include the use of 'visual variables' (Bertin 1967), 'image schemata' (Johnson 1987; Lakoff 1987) and Colin Ware's (2008) 'graphical codings' (see Table 1). In this context we propose to distinguish nine different types of information that can be encoded visually, together with the kind of questions that are answered by these types of information. See Fig. 1.

- **Depicting** is used to represent aspects of the *visual appearance* and/or *spatial location* of entities in the physical world (existing or imagined). This type of information answers questions such as *What does it look like?* and *Where is it?* Depictions – pictures as well as maps – may range in a spectrum from realistic/precise to schematic, which we discuss later in the section titled '*Mode of depiction*'.
  - **Picturing** can answer *What does it look like?* as well as *Where is it?* questions. This can be done by representational methods such as perspective projection. *Picturing* is the key visual encoding principle used in technical illustrations and other pictorial representations.

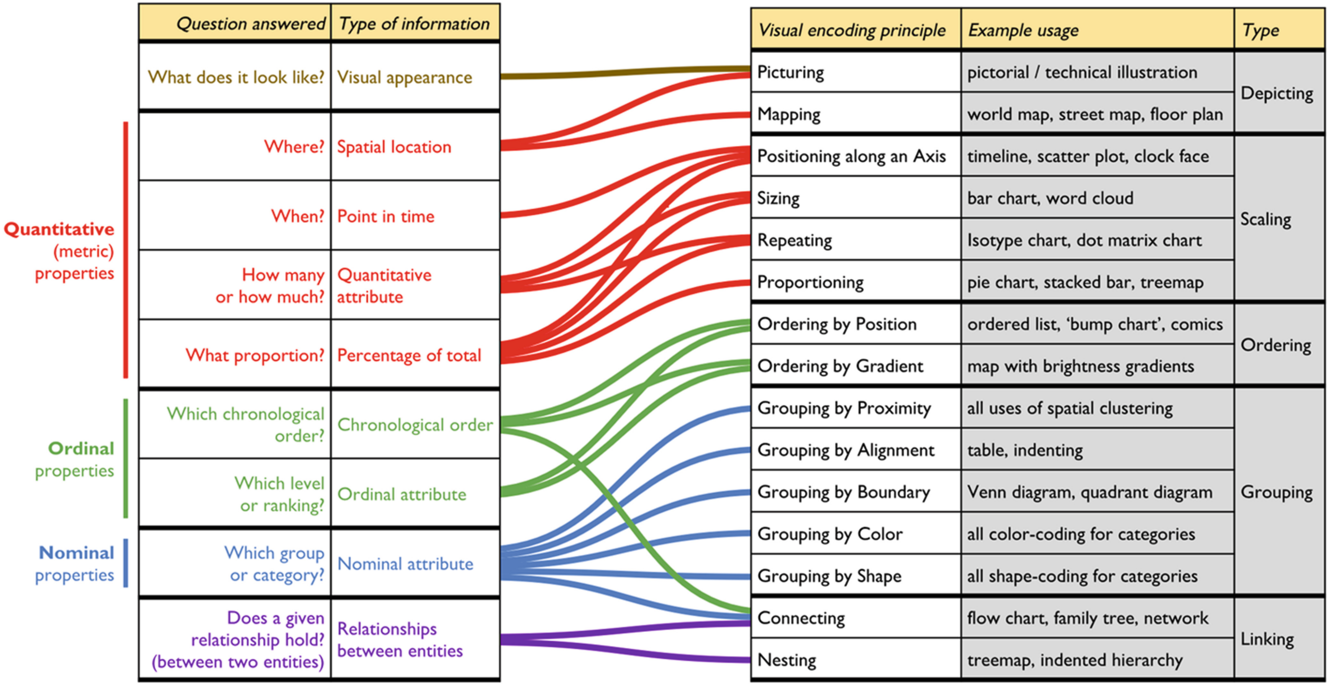


**Fig. 1.** Interdependencies between *types of information* and the *visual encoding principles* that can be used to represent them, together with example usages and the five visual encoding *types*.

  - **Mapping** is used to display the two-dimensional layout of physical configurations (*spatial location*). *Mapping* answers *Where?* questions. This can be done by representational methods such as cartographic projection. *Mapping* is the key visual encoding principle in world maps, street maps and floor plans.
- **Scaling** is used to represent *quantitative attributes* of entities (e.g. amounts, numerical values), or *percentage of total*, or *point in time. Scaling* can answer questions such as *How many?*, *How much?*, *What proportion?* or *When?*
  - **Positioning along an axis** can be used to represent and answer all of the above, and is the key visual encoding principle in timelines, clock faces, line charts and scatter plots. An axis can be embedded in rectangular coordinates (horizontal, vertical) or in polar coordinates (angular, radial).
  - **Sizing** is used to represent *quantitative attributes* or *percentage of total* and is the key visual encoding principle in bar charts and word clouds.
  - **Repeating** is the use of multiples of graphic components arranged into arrays of proportional size. Like *sizing*, *repeating* is used to represent *quantitative attributes* or *percentage of total. Repeating is* the key visual encoding principle in Isotype charts and dot matrix charts.
  - **Proportioning** is used to represent *percentage of total* and answers *What proportion?* questions. *Proportioning* is done by dividing a given surface area into proportional segments and is the key visual encoding principle in pie charts, stacked bars and 'treemaps'.
- **Ordering** is used either to represent *ordinal attributes* of entities, such as level or rank, or to represent *chronological order. Ordering* can answer questions such as *Which level?*, *Which ranking?* or *Which chronological order?*
  - **Ordering by position** is a key visual encoding principle in comic strips, pictorial assembly instructions, ordered lists and 'bump charts'.
  - **Ordering by gradient** is the key visual encoding principle in maps with grayscale or other brightness gradients and in tabular heatmaps.
- **Grouping** is used to represent *nominal attributes* of entities, i.e. category membership. *Grouping* answers questions such as *Which group?* and *Which category?*
  - **Grouping by proximity** is a visual encoding principle in all information graphics that use spatial clustering.
  - **Grouping by alignment** is a key visual encoding principle in indented lists and tabular representations.
  - **Grouping by boundary** can be done by separating graphic components by a demarcating line, enclosure or shared background. *Grouping by boundary* is the key visual encoding principle in quadrant diagrams and Venn diagrams.
  - **Grouping by color** is a visual encoding principle in all information graphics that use color-coding to identify categories.
  - **Grouping by shape** is a visual encoding principle in all information graphics that use shapes to identify categories.
- **Linking** is used to represent the presence or absence of *relationships between entities*, such as connections, pathways, chronological order or hierarchies. *Linking* answers the question *Does a given relationship hold (between two entities)?* (e.g. 'is a friend of', 'is a part of').

- **Connecting** graphic components can be done by other graphic components that function as joining devices such as lines or arrows. *Connecting* is the key visual encoding principle in flow charts, family trees and network graphs. The relationships between entities that can be represented by *connecting* include *nominal attributes* (relationship: ‘is in the same category as’) and *chronological order* (relationship: ‘precedes’).
- **Nesting** is usually used in combination with either *grouping* or *connecting*.
  (a) In combination with *grouping*, *nesting* can be done by grouping a subset of graphic components (by *proximity* or *boundary*) which together form a component within a higher-level group. For example, *grouping by boundary*, *nesting* and *proportioning* are the three key visual encoding principles in ‘treemap’ and ‘circle packing’ visualizations. In combination with *grouping by alignment*, *nesting* can be done by indenting a graphic component below another graphic component. These are key visual encoding principles in indented lists.
  (b) In combination with *connecting*, *nesting* is a key visual encoding principle in tree structures. Tree structures achieve *nesting* by combining *connecting*, *ordering by position* (usually vertically) and often also *grouping by alignment* (usually horizontal alignment).

This may seem to be a limited set of *visual encoding principles*, however, these can be used in an enormous range of combinations. The framework can thus offer a potential tool for creating alternative representations. For any given visualization or data set a decision-tree can be followed (along the colored lines in Fig. 1) from information types to possible choices of *visual encodings*, resulting in a more diverse

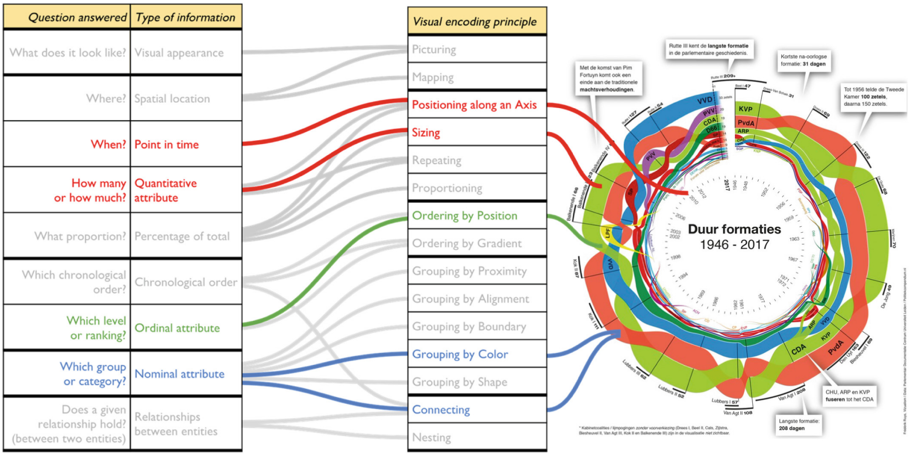


**Fig. 2.** On the right: diagram showing the number of seats per political party in Dutch parliament for the last 71 years, displayed on an angular time axis. The color-coded political parties are ordered from smallest on the inside to largest on the outside. (Visualization by Frédérik Ruys.) On the left: types of information that are represented and *visual encoding principles* that are used.

range of combinatorial possibilities than provided by other frameworks (Richards and Engelhardt forthcoming). The combining of these visual encoding principles is part of what makes visual representations so powerful. One combinatory possibility is shown in Fig. 2.

## 5 Mode of Correspondence

Each *visual encoding* can be characterized by its *mode of correspondence*. The *mode of correspondence* refers to how the *graphic relationships* displayed may correspond to the relationships that are represented. A visual encoding may involve either a *literal* (physical) or a *non-literal* mode of correspondence (e.g. by metaphor or convention).

**Literal** visual encodings display graphic relationships that *show a degree of structural similarity to actual physical relationships that they represent* – **or** they display graphic qualities or objects that *show a degree of resemblance to actual visual appearances of physical objects (existing or imagined) that they represent*. (Also see Table 1.)

**Non-literal** visual encodings display graphic relationships that *do not represent* actual physical relationships, *but rather conceptual relationships* – **or** they display graphic qualities or objects that *stand for what is meant by metaphor or convention*.

In the case of *connecting* for example, the lines in a wiring diagram show physical connections, thus involving *literal* correspondence, whereas the lines in a family tree represent links between family members metaphorically, thus involving *non-literal* correspondence.

## 6 Mode of Depiction

The *mode of depiction* describes the possibilities regarding *the type of visual encoding* identified as *depicting* (*picturing* and *mapping*). The *mode of depiction* ranges from the **realistic/precise** (e.g. spatially or visually realistic and/or precise) to the **schematic** (e.g. spatially or visually synoptic and/or edited).

In the case of *picturing*, a detailed realistic drawing is an example of a *realistic/precise* mode of depiction, while a stick figure or a smiley face are examples of a *schematic* mode of depiction. In the case of *mapping*, a detailed topographic map of a mountain range is an example of a *realistic/precise* mode of depiction, while a subway map or a cartogram are examples of a *schematic* mode of depiction.

If (a component of) a visual representation involves neither picturing nor mapping, than we refer to it as **non-depictive.**

## 7 Conclusions

Based on a thorough analysis of previous frameworks and of specimens from two corpuses of visualization types, we conclude that the framework presented here provides a comprehensive taxonomy for analyzing, codifying and comparing the key distinguishing features of diagrams and graphics. It offers a potential method for identifying unexplored combinations of visual encoding principles, and for creating alternative representations for any given visualization or data set.

## References

Bertin, J.: Semiology of Graphics: Diagrams, Networks, Maps. University of Wisconsin Press, Madison (1967/1983). (original: Sémiologie graphique. Gauthier-Villars)

Bertin, J.: Graphics and Graphic Information Processing. Walter de Gruyter & Co, Berlin (1977/1981). (original: Le graphique et le traitement graphique de l'information. Flammarion.)

Blackwell, A., Engelhardt, Y.: A meta-taxonomy for diagram research. In: Anderson, M., Meyer, B., Olivier, P. (eds.) Diagrammatic Representation and Reasoning, pp. 47–64. Springer, London (2002). https://doi.org/10.1007/978-1-4471-0109-3_3

Card, S.K., Mackinlay, J.: The structure of the information visualization design space. In: Proceedings of the 1997 IEEE Symposium on Information Visualization (InfoVis 1997), pp. 92–99. IEEE Computer Society, Washington, DC (1997)

Engelhardt, Y.: The language of graphics. Ph.D. thesis, University of Amsterdam (2002)

Johnson, M.: The Body in the Mind. University of Chicago Press, Chicago (1987)

Lakoff, G.: Women, Fire, and Dangerous Things. The University of Chicago Press, Chicago (1987)

Maxwell, J.C.: Diagrams. In: Encyclopaedia Britannica, 11th edn. (1910)

Peirce, C.S.: The Collected Papers of Charles Sanders Peirce, 4.418. Harvard University Press, Cambridge (1903/1933)

Richards, C.J.: Diagrammatics. Ph.D. thesis, Royal College of Art, London (1984)

Richards, C.J., Engelhardt, Y.: Diagrammatics: the fundamental modes and building blocks of visual representations. In: Richards, C.J. (ed.) Elements of Diagramming: Theoretical Frameworks, Design Methods, Domains of Practice. Routledge, London (forthcoming)

Russell, B.: Vagueness. Australas. J. Psychol. Philos. **1**, 84–92 (1923)

Tversky, B.: Cognitive origins of graphic productions. In: Marchese, F.T. (ed.) Understanding images, pp. 29–53. Springer, New York (1995)

Twyman, M.: A schema for the study of graphic language. In: Kolers, P.A., Wrolstad, M.E., Bouma, H. (eds.) Processing of Visible Language, vol. 1, pp. 117–150. Plenum Press, New York (1979)

Ware, C.: Visual Thinking for Design. Morgan Kaufmann Publishers, San Francisco (2008)

# A Classification of Infographics

Helen C. Purchase[1](✉), Katherine Isaacs[2], Thomas Bueti[2],
Ben Hastings[2], Aadam Kassam[2], Allen Kim[2],
and Steffan van Hoesen[2]

[1] School of Computing Science, University of Glasgow, Glasgow, UK
`helen.purchase@glasgow.ac.uk`
[2] Department of Computer Science, University of Arizona, Tucson, USA

**Abstract.** Classifications are useful for describing existing phenomena and guiding further investigation. Several classifications of diagrams have been proposed, typically based on analytical rather than empirical methodologies. A notable exception is the work of Lohse and his colleagues, published in Communications of the ACM in December 1994. The classification of diagrams that Lohse proposed was derived from bottom-up grouping data collected from sixteen participants and based on 60 diagrams. Mean values on ten Likert-scales were used to predict diagram class. We follow a similar methodology to Lohse, using real-world infographics (i.e. embellished data charts) as our stimuli. We propose a structural classification of infographics, and determine whether infographics class can be predicted from values on Likert scales.

**Keywords:** Infographics · Classification · Empirical studies

## 1 Introduction

Infographics present quantitative data (like that in bar charts or scatterplots), and are typically embellished with graphic elements or pictures. Infographics can increasingly be found in popular media, online, in public presentations and organisations' brochures, making data more visible, engaging, and memorable. Several researchers investigate the effect of using embellishments in data presentation by conducting empirical studies, the stimuli sometimes "real" (sourced from media publications) and sometimes "fabricated" (created by researchers for the purposes of their experiment).

With increasing infographics research, classification is useful. "A carefully designed classification can serve to show not only the full range of available possibilities but also the relationships between these, and … acts more as an instrument rather than simply as a 'filing cabinet'" (Rankin [1]). Kwasnik [2] explores the relationship between classifications and knowledge discovery: "Classification is a way of seeing. Phenomena of interest are represented in a context of relationships that, at their best, function as theories by providing description, explanation, prediction, heuristics, and the generation of new questions."

Classifications can be generated by thorough and systematic analysis of a range of stimuli [3–5], or by soliciting the views of human participants (Lohse et al. [6]). The

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 210–218, 2018.
https://doi.org/10.1007/978-3-319-91376-6_21

research reported in this paper takes the latter approach: we conducted an empirical study to create a classification of infographics, based on “real” stimuli.

## 2 Prior Research

Garcia and Cox [4] considered diagrams in the UK National School Curriculum, classifying them into 20 types, and discussing them with respect to children’s “graphical readiness” to interpret diagrams. Purchase [3] analysed diagrams from the proceedings of the first seven conferences on the Theory and Application of Diagrams: her primary classification is abstract vs concrete and embellishments are defined as ‘additional visual elements’. Novic [7] classes scientific diagrams as “iconic”, “charts and graphs” and “schematic diagrams”. Blackwell and Engelhardt [8] surveyed several diagram taxonomies, noting differences according to the nature of the visual elements used, their positioning, their semantics, and context of use. Rankin [1] commented on the diversity of classification criteria used by different researchers, distinguishing between two types of diagrammatic classification: functional (focusing on purpose) and structural (focusing on form). Our motivator is the CACM article by Lohse et al. [6], who presented the first structural classification of diagrams based on empirical data, collected from 16 participants.

The term ‘infographic’ is defined in many different ways. Saleh et al. [11] write: “Infographics are complex graphic designs integrating text, images, charts and sketches”. Albers [5] writes: “an infographic takes a large amount of information in text or numerical form and condenses it into a combination of text images and with a goal of making the information presentable.” We wished to focus on the metaphorical use of graphical elements (e.g. pictures of coins, cakes, monkeys, suitcases, wine glasses) as a means of depicting data: that is, if these graphical elements were removed from the image, then this would remove the representation of the data. So, a bar chart with a picture of the moon in the background is not an infographic; a bar chart where each bar is represented by a picture of a space shuttle of a different height is. Haroz et al. [12] discovered that superfluous images not used for representing data were distracting, and so we insist that any graphics items directly depict data values.

Albers [5] used an ‘open-ended card sort’ method on 25 infographics to devise four categories: bullet list equivalent, snapshot with graphic needs, flat information with graphic needs, and information flow/process – a categorisation formed from the author’s personal view. Borkin et al. [10] do not describe how they created the 12 categories in their ‘visualisation taxonomy’; Saleh et al. [11] investigate the ‘stylistic similarity’ of infographics, but do not explicitly identify or name different ‘styles’. Popular websites (e.g. excelcharts [13], juiceanalytics [14]) propose classifications of data charts, but do not include charts with graphical embellishments.

# 3 Methodology

We follow the empirical and data analysis methodology of Lohse et al. [6] closely, our objectives being to create a hierarchical taxonomy of different types of infographics, and devise a means of predicting the class of an infographic, based on the responses to ten Likert scales. Our empirically-derived classification structure can inform further empirical research on infographics.

We use the following ‘infographic’ definition: “An image that presents a data set, where the data quantities are depicted using pictures of recognisable common items.”

## 3.1 Materials

We used existing data sets: Saleh et al.’s [11] set of 19,594 infographics, and Borkin et al.’s [9] 5,693. In addition, we looked at 55 infographics from the Times Higher Education magazine and a set of 174 infographics previously gathered from a range of sources. Most images were eliminated quickly because they presented more than one data set, were of poor resolution, were duplicates, had an extreme aspect ratio, had text not in English, were photographs, or were data charts not embellished with images. We eliminated those where the images or pictures used to embellish the data chart were not integral to the presentation of data. We chose 60 infographics to ensure data presentation method variety (see www.dcs.gla.ac.uk/~hcp/infographics).

Our starting point for devising our Likert scales was Lohse et al.’s original ten [6], although we also drew from those used by Quispel [15], Loroco et al. [16], and Harrison et al. [17]. Our scales are: spatial/non-spatial; non-temporal/temporal; hard to understand/easy to understand; concrete/abstract; attractive/unattractive; emphasizes the whole/emphasizes parts; informative/uninformative; minimal/cluttered; shows patterns/does not show patterns; literal/metaphorical.

## 3.2 Experimental Procedure

Twenty participants took part (10 female, mean age = 33, 9 students, 3 high school graduates, 8 university graduates). Three were studying computer science, and the rest were a mixture of a variety of subjects (e.g. Law, Social Work, Business); none were studying visualization, graphic design or art. Each experiment was conducted one-on-one, and took approximately 90 min.

Table 1 shows how our procedure differs from that of Lohse et al. [6]. Each participant was given the 60 infographics in a pile, in a different random order for each participant, and asked to describe briefly, aloud, what each infographic was about. They then laid all the infographics out on the table and grouped them according to “visual design.” If participants were not sure what was meant by the phrase “visual design”, this was explained to them using phrases like “the way in which the graphic has been designed”, or “the overall visual design of the infographic.” They could have as many groups as they liked, as many infographics in each group as they liked, and could take as long as they wished. They then explained their rationale behind each group. After a break, the participants rated each infographic on the ten Likert scales.

**Table 1.** Experimental procedure.

| | Lohse et al. [6] | Our experiment |
|---|---|---|
| Stimuli | 60 diagrams, chosen to be "representative…within the domain of static, two-dimensional graphic representations" | Infographics with primary aim of presenting quantitative data, embellished with images |
| Familiarisation | Participants named each diagram (step 1) | Participants described what each infographic is "about" (step 1) |
| Rating | Participants rated each diagram on ten nine-point Likert scales (step 2) | Participants rated each diagram on ten nine-point Likert scales (step 4) |
| Grouping | Participants performed a bottom-up sorting task on randomly laid out diagrams, grouping items with respect to "similarity" (step 3) | Participants performed a bottom-up sorting task on randomly laid out diagrams, grouping items with respect to "visual design" (step 2) |
| Explanation | Participants gave the rationale for their grouping (step 4) | Participants gave the rationale for their grouping (step 3) |

### 3.3 Data Analysis

We follow the data analysis procedure of Lohse et al.

(1) Outlier pruning. We calculate the distance between pairs of participants using Jaccard coefficients: the distance between participants Pi and Pj is $1 - A/(N - B)$ where N is the number of infographic pairs ($60 * 59/2 = 1770$), A is the number of infographic pairs that appear together in both Pi and Pj's groupings, and B is the number of infographic pairs that appear in separate groups in both Pi and Pj's groupings. Complete linkage hierarchical clustering on the matrix of Jaccard coefficients produced a tree: participants on singleton branches until final mergings are considered outliers.

(2) Classification of Infographics. We derive a hierarchical cluster tree of infographics using complete linkage hierarchical clustering. The similarity matrix comprises similarity scores for infographics pairs: the number of participants who put the pair in the same group. We normalized the similarity and subtracted from one to convert to distances for clustering. The existence of ties in distance scores leads to different hierarchical clusterings based on the ordering of infographics in the matrix. Following Lohse et al., we computed six hierarchical clusterings, permuting the matrix each time.

(3) Predicting the classification. We use average Likert scores for each infographic. We perform a principal components analysis (PCA) on the rating scales to determine if any scales should be removed due to explaining little of the variance. With the remaining scales, we then build two classifiers, one using classification and regression trees (CART) and one using linear discriminant analysis (LDA).

Per the requirements of the scikit-learn library and following Lohse et al., we input cluster priors through the ‘class_weight’ parameter for CART and as a passed parameter for the LDA. Both the CART and LDA were evaluated using 11-fold cross-validation (as in Lohse et al.). We used the default Gini index as the splitting criterion for the CART analysis.

## 4 Results

Outliers are participants on singleton branches until the final stages of merging; our clustering yielded one such participant who grouped by subject matter rather than design. Further analysis of the reasons participants gave for their grouping indicated three others focused on attributes other than ‘visual design’ (e.g. colour, semantics, audience). We removed these four participants’ data, leaving 16 valid data sets.

We set the similarity distance threshold to 0.9, resulting in seven to eight clusters for each of the six cluster trees. We inspected these six clusterings to form a meta-clustering by grouping infographics that appeared in the same cluster in the majority. Our classification analysis revealed six top-level categories, two of which are comprised of two second-level sub-categories. Two infographics appeared with similar frequency in two categories (in the ‘area-as-quantity’ and ‘single circle’ classes): they both presented two sets of data. We had attempted to ensure that each infographic only presented one data set – these two had slipped through the net of our filtering process so were removed from further analysis. Two other multiply-classified infographics were both based on flags – we therefore created a separate ‘flag’ category for them, the seventh top-level category. The seven categories are:

- **Bar Charts** (16). A bar chart is the main data presentation form.
- **Geographical** (4). The primary shape is a geographical map.
- **Units** (6). The quantity of the data is represented by several small graphic images, each representing an amount of data.
- **Area-as-Quantity**. Different data quantities are represented by the areas of shapes. In some cases, these are **Familiar Shapes** (e.g. circles, triangles) (9); in others **Uncommon Shapes** are used (e.g. dinosaurs, mail boxes) (5).
- **Single Circle** (5). Data is represented within a singular circular form.
- **Proportion-as-Quantity**. The data quantities are shown as proportions of a larger object. Divisions of **Rectangular Shapes** are most common (6), although **Irregular Shapes** (e.g. banana, wine glass) are also used (5).
- **Flags** (2). The primary shape used is that of national flags.

The first three principal components accounted for 91.1% of the variance. Each Likert scale had a squared factor loading >10% in at least one of the first three principal components. Thus, we chose to keep all of the scales. To avoid overfitting the CART tree, we set the maximum number of leaves to 10, similar in detail to Lohse et al. [6]. The resulting tree correctly classifies 55.2% of the infographics with a cross validation mean accuracy of 28%. Examining the CART tree and the distribution of average Likert values for all of the infographics, we observed there is a high degree of variance within many of the clusters for each Likert. For example, paired bar charts often represent before and after, giving them a higher temporal score than the non-paired bar charts in the bar chart group. The LDA resulted in a slightly more accurate classifier (63.8%, with cross validation mean accuracy of 38%).

## 5 Discussion

Some specific infographics produced surprising results. A line chart (i03, see website in Sect. 3.1) was consistently grouped as a bar chart; its source was The Times Higher magazine, as was the case for several bar charts – perhaps there is a common generic 'Times Higher' visual style that led it to be grouped with others from the same source? Alternatively, since this was the only infographic based on a 'line chart', it may have been grouped with bar charts so as to not be a singleton group. The cartogram (i19) was the extreme on several Likert scales, and was not classified as 'Geographic'. We believe that some participants did not recognise it as representing a world map. An infographic which represented money as piles of poker chips (i57) was not classified as a bar chart; however, since the individual piles of chips have no meaning, and it is the comparative area of the two piles that is important, 'Area-as-Quantity' is indeed the best classification for it. The map of Africa showing how its area compared to that of other countries (i23) was predominantly classified as Geographical, although it might also reasonably be in the Proportion-as-Quantity (Irregular Shapes) or Area-as-Quantity (Uncommon Shapes) categories. i58 might have better been classified in the Area-as-Quantity class (Familiar Shapes) – we believe that the highly rectangular nature of the items depicted in it led it to be grouped with the other Rectangular Shapes as part of Proportion-as-Quantity. We deliberately included an infographic that depicted a single data point (i26) as an extreme example; it was classified as Proportion-as-Quantity since, we believe, the range was implicitly interpreted as [−40°F, 140°F], the common range of thermometers of that design.

Our classification is richer than those of Albers [5] and Borkin et al. [10], which are based on popular categories of data charts (e.g. donut chart, stacked area chart, line, scatter plot, tree [10]) or are vague (e.g. "flat information with graphic needs" [5]). Some of our empirically-derived classifications are similar to common data charts (e.g. Bar Charts, Geographical), but they also include categories based on how the space on the page is used to depict data (e.g. Proportion-as-Quantity). Unlike other classifications, our results show that participants were not only aware of how data was being depicted (e.g. using proportions to show quantity), but were also highly sensitive to the types of shapes used – familiar, uncommon, rectangular, irregular. No other classification considers the form or shape of the graphical embellishments used.

There is a strong prevalence of infographics that rely on area comparisons to show difference in data values: 14 Area-as-Quantity and 11 Proportion-as-Quantity. It is well known that perception of area is less accurate than perception of length or position [18]. This phenomenon might actually serve infographics designers who wish mislead readers: Tufte [19, pp. 69–70] gives examples where perception of area rather than length can easily lead to incorrect inferences.

The Likert scales were poor predictors of class, in contrast to Lohse et al's results. The data indicates that the Likert scales are orthogonal to the classifications – that is, their values bear little relation to the groupings created by the participants. Thus, whether an infographic is attractive or not, or easy to understand, or temporal etc. does not reflect its visual form. In many ways, this is reassuring news for infographics designers – they are not obliged to use any of the nine specified categories if they wish to emphasise any of these Likert properties. In addition, Lohse et al. suggested that their successful predictions might have been a result of participants doing the Likert scales before the grouping task, and then implicitly using these scales in their grouping. Our participants completed the Likert scales after the grouping task, so as to mitigate against this possibility. Having the two tasks done by different participants of similar demographic profile might be a more reliable way of testing the predictions: that way, there would be no cross-contamination between tasks.

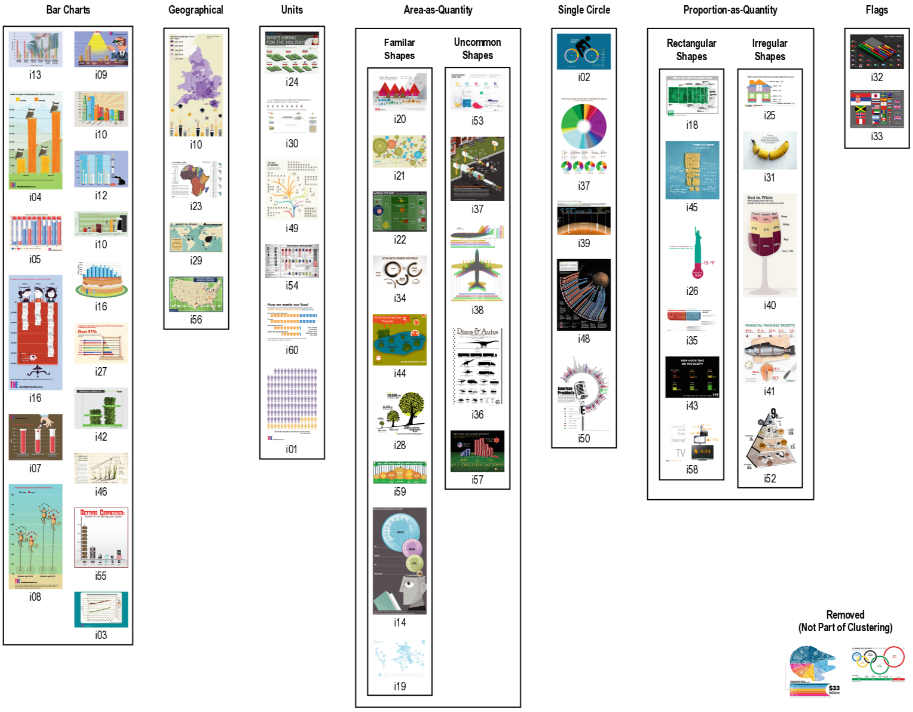
Any empirical study is subject to limitations. Our classification results are bound by the scope of the 60 infographics we chose (from a total set of 25,516), and our prediction results by our choice of Likert Scales. The demographic profile of our participants is reasonably well-spread, although slightly skewed towards younger ages. Future work can validate our hierarchy with other infographics and participants.

## 6 Conclusion

The prevalence of infographics in the popular media, advertising, public notices and organizational brochures makes them a rich source for diagrammatic research. There is still a great deal of empirical work to be done in this area: what makes infographics memorable or engaging? Do graphical embellishments inhibit interpretation – both of individual data points or the overall message? How can deliberately misleading messages be presented without being obvious? Classifications provide frameworks for research, and are particularly useful if based on real-world examples and created through human experimentation. Our novel classification of infographics provides an empirically derived basis for researchers in this area – who no longer need to create their own analytical classifications.

**Acknowledgements.** This research was conducted while the first author was visiting the University of Arizona. We are grateful to all the participants, and to Babak Saleh and Michelle Borkin who shared their image data sets. Ethical approval was given by the University of Arizona Institutional Review Board (ref: 1711982836).

## Appendix: The Infographics Classification Tree



## References

1. Rankin, R.A.: A taxonomy of graph types. Inf. Des. J. **6**(2), 147–159 (1990)
2. Kwasnik, B.: The role of classification in knowledge representation and discovery. Libr. Trends **48**, 22–47 (1999)
3. Purchase, H.C.: Twelve years of diagrams research. J. Vis. Lang. Comput. **25**, 57–75 (2014)
4. Garcia, G.G., Cox, R.: Diagrams in the UK national curriculum. In: Stapleton, G., Howse, J., Lee, J. (eds.) Proceedings of the Fifth International Conference on the Theory & Application of Diagrams. Lecture Notes in Computer Science, LNAI, vol. 5223. pp. 360–363. Springer, Berlin (2008). https://doi.org/10.1007/978-3-540-87730-1_36
5. Albers, M.J.: Infographics: Horrid chartjunk or quality communication. In: IPCC, pp. 13–15. IEEE (2014)
6. Lohse, G.L., Biolsi, K., Rueter, H.H.: A classification of visual representations. Commun. ACM **37**(12), 36–49 (1994)
7. Novick, L.R.: The importance of both diagrammatic conventions and domain-specific knowledge for diagram literacy in science: the hierarchy as an illustrative case. In: Barker-Plummer, D., Cox, R., Swoboda, N. (eds.) Diagrams 2006. LNCS (LNAI), vol. 4045, pp. 1–11. Springer, Heidelberg (2006). https://doi.org/10.1007/11783183_1

8. Blackwell, A.F., Engelhardt, Y.: A meta-taxonomy for diagram research. In: Anderson, M., Meyer, B., Olivier, P. (eds.) Diagrammatic Representation and Reasoning, pp. 47–64. Springer, London (2002). https://doi.org/10.1007/978-1-4471-0109-3_3
9. Borkin, M.A., Bylinskii, Z., Kim, N.W., Bainbridge, C.M., Yeh, C.S., Borkin, D., Pfister, H., Oliva, A.: Beyond memorability: visualization recognition and recall. Vis. Comput. Graph. IEEE Trans. **22**(1), 519–528 (2016)
10. Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., Pfister, H.: What makes a visualization memorable? IEEE Trans. Vis. Comput. Graph. **19**(12), 2306–2315 (2013)
11. Saleh, B., Dontcheva, M., Hertzmann, A., Liu, Z.: Learning style similarity for searching infographics. In: Proceedings of the 41st Graphics Interface Conference (GI 2015), pp. 59–64. Canadian Information Processing Society, Toronto (2015)
12. Haroz, S., Kosara, R., Franconeri, S.L.: ISOTYPE visualization: working memory, performance, and engagement with pictographs. In: Proceedings CHI 2015, pp. 1191–1200. ACM, New York (2015)
13. A classification of chart types, https://excelcharts.com/classification-chart-types/. Accessed 15 Nov 2017
14. Juice Analytics Chart Chooser, http://labs.juiceanalytics.com/chartchooser/index.html. Accessed 15 Nov 2017
15. Quispel, A.: Data for all: how professionals and non-professionals in design use and evaluate information visualizations. Ph.D. thesis, Tilburg University (2016)
16. Locoro, A., Cabitza, F., Actis-Grosso, R., Batini, C.: Static and interactive infographics in daily tasks: a value-in use and quality of interaction user study. Comput. Hum. Behav. **71**, 240–257 (2017)
17. Harrison, L., Reinecke, K., Chang, R.: Infographic aesthetics: designing for the first impression. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, pp. 1187–1190. ACM, New York (2015)
18. Cleveland, W.S., McGill, R.: Graphical perception: theory, experimentation, and application to the development of graphical methods. J. Am. Stat. Assoc. **79**(387), 531–554 (1984)
19. Tufte, E.R.: The Visual Display of Quantitative Information. Graphics Press, Cheshire (1983)

# Diagrammatic Maps of the New York Subway: An Historical Perspective

Peter B. Lloyd(✉)

School of Computing, University of Kent,
Darwin Road, Canterbury CT2 7NZ, UK
pbl@kent.ac.uk

**Abstract.** Vignelli's 1972 diagrammatic subway map is hailed as a design classic, but was dropped by the Metropolitan Transportation Authority (MTA) after just seven years' usage. Following an absence of a generation, a diagrammatic map of the New York City subway system has been reintroduced into the MTA's information provision. A digital version came back in 2011 and continues in use with weekly updates on the MTA Weekender website; print editions were issued in 2012, 2014, 2015, and 2017 for special occasions and from 2017 onwards for travel advisory notices. To see this in context, we need to understand why New York City adopted a diagrammatic map (Salomon map 1958), route colour-coded it (D'Adamo map 1967), stylized it (Vignelli map 1972), replaced it with a geographic map (Tauranac map 1979), and re-imagined it for the digital era (Waterhouse-Cifuentes map 2011). Using primary sources, we characterise the birth, death, and rebirth of the diagrammatic map of the New York City subway.

**Keywords:** Metro maps · New York City subway · Salomon
Vignelli · MTA

## 1 Introduction

The two most famous metro maps are probably the London Underground map by Henry C. Beck (1933, with derived versions continuing today), and the New York City subway map by Massimo Vignelli (1972 to 1978). The brevity of the Vignelli map in contrast with the longevity of the Beck map begs explanation, as does the reimagining of the Vignelli map thirty-two years later by Waterhouse and Cifuentes (Lloyd 2012). Against a worldwide trend for metro maps to be diagrammatic (Ovenden 2015), the adoption of a geographic design as the official subway map of the MTA (Metropolitan Transportation Authority) for more than three decades stands out as an anomaly. In this paper, we address the question of why it underwent these several transitions. This paper is based on information collected from primary sources during the period 2003 to 2018: acquisition of publicly issued maps; face-to-face interviews with surviving individuals involved in the main transitions of the official subway map; inspection of contemporary manuscript and typescript documents, and contemporary news reports.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 219–227, 2018.
https://doi.org/10.1007/978-3-319-91376-6_22

**Designation of the Maps.** Since the New York Transit Authority (TA)'s first in-house map, in 1958, the official subway map has been anonymous—except for 1998 to 2009, when Michael Hertz put his design firm's name on it. Before the Salomon map of 1958, the municipal authority anonymized its map, while private firms indicated individual authorship: the IRT printed the initials of its map's designers ("HLS" and "JWG"); the BMT printed the name of its map's designer ("G. V. Plachy"), the Hagstrom maps—which were adopted by the Board of Transportation (BoT) and the TA—had the cartographic firm's name ("Hagstrom Map Co."), and the Voorhies map adopted by the TA had "Stephen Voorhies". Formally, the TA and from 1972 its parent the MTA was 'the designer' of the subway map. For the purposes of this paper, however, official designs of the subway map will be referred to by the individual who instigated the distinguishing features. This is a convenient label rather than an ascription of an *auteur* to a map. So, 'the Salomon map' refers to editions of the official map of the New York City subway from Winter 1958 to Spring 1967; 'the D'Adamo map' refers to editions from Autumn 1967 to 1969 inclusive; 'the Vignelli map' refers to those from August 1972 to 1978 inclusive; 'the Tauranac map' refers to editions from June 1979 to 2011, and that series continued to the present times; and 'the Waterhouse-Cifuentes map' refers to the Weekender map and print renderings thereof.

Large changes of a map herald a new *design*, while the smaller increments introduce *versions* and *variants*, although Roberts (2018) advocates a strict notion of design succession, in which any non-trivial changes constitutes a new design. Here we deem that the New York City subway shifted to new *designs* in October 1958 (Salomon), November 1967 (D'Adamo), August 1972 (Vignelli), June 1979 (Tauranac), and September 2011 (Waterhouse-Cifuentes), and other editions are deemed to be *versions*.

## 2 Findings

**Before the First Diagrammatic Map.** The subway in New York City was built by three companies (IRT, BMT, and IND), who ran their networks independently until they were unified under municipal control (the BoT) in the summer of 1940. Until then, each company made a map showing only its own services. Private cartographers (such as Hagstrom, Nostrand, Voorhies, General Drafting) made maps of the complete network, which were either sold to passengers through newsagents, or overprinted with promotional material and circulated free of charge by hotels, banks, conventions and other businesses. After unification under the BoT in 1940, almost two decades passed before they commissioned a new map: black-and-white versions of the former operating companies' maps continued to be issued by the BoT for a few years. Then they started issuing Hagstrom's map of the integrated system, and the TA continued this practice from 1953—alongside, from 1954, Voorhies' map (overprinted the Union Dime Bank's promotional material). These topographic maps continued until the autumn of 1958.

**1958: The Salomon Map—The First Diagram.** George Salomon (1920–1981) came as an émigré via London and in 1940 settled in New York City. He was inspired by London Underground's signage and map to create a systematic service nomenclature, colour-coding scheme, signage system, and network map for the New York City subway (Salomon 2006, Salomon 2003). His nomenclature also resembled the trunk-and-branch scheme of the early 1950s Berlin U-Bahn map which he probably received copies of. Salomon aligned himself with the Bauhaus school, and with modernist artists, especially Mondrian, who settled in New York in the same year (Salomon 2004).

By 1948, Salomon was actively working on his concepts for way finding in the subway, which was at first a private project (Salomon 2006). In 1953, the TA succeeded the BoT, bringing a proactive approach to promotion and information delivery. Salomon approached the TA immediately (Salomon 1956b), and by 1956 had submitted two prospectuses outlining his concept for an overhaul of wayfinding: renaming the routes and colour-coding them using a trunk-and-branch structure, systematising the signage, and creating a diagrammatic map (Salomon 1955, 1956a). What he proposed to them was the culmination of several years of his personal research. In September 1956, the TA selected Salomon's map (Salomon 1956b), but kept the nomenclature and tricolour scheme of the long-gone IRT, BMT, IND. They issued their first diagrammatic pocket map, designed by Salomon, in October 1958 (Fig. 1), which also appeared in carriages and on station walls over the months from December onwards. The TA had commissioned the map to solve problems with outsourced maps, which had been apparent under the BoT (Daly 1952) and would be exacerbated by increased print runs from about 50,000 maps (at $32.50 per thousand) a year (BoT 1952) to about 500,000 a year. The use of Union Dime's free Voorhies maps from 1954 could solve the cost problem, but the map leaflet was dominated by Union Dime rather than the subway body. Moreover, the lack of direct control over editorial content and the slowness of updates remained problematic. Salomon's overtures, motivated by his passion for better wayfinding, converged with the TA's desire for cheaper and easier in-house mapmaking. Internal memoranda reveal this as the actual motivation, while public-facing documents indicate a *post hoc* rationalization: "A new subway map has been designed to simplify the problems of those who seek to find their way around the city on rapid transit lines," from the Annual Report a few months before the map was launched. (TA 1958a; see also TA 1958b and TA 1959).

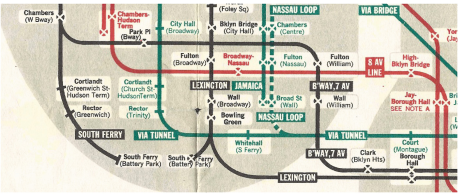
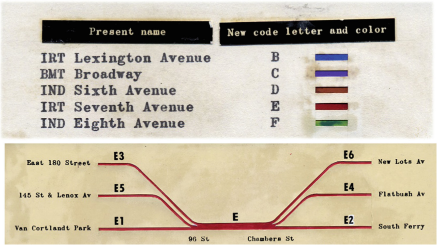


**Fig. 1.** Excerpt from Salomon 1958 map

**1958: Colours.** There was no prior official colour coding: from 1943, the BoT had used the Hagstrom map with spot red (IRT), process blue (BMT), and process orange (IND), although the latter had changed to spot yellow by 1948. From 1953, the TA continued using that map and, from 1954, added the Voorhies map, which had spot blue (IRT), process orange (BMT), and spot red (IND). So, when Salomon prepared his report, he had no official constraint on possible colour schemes. He proposed a nomenclature that he based on the main trunks running north-south in Manhattan, giving each trunk a letter code and a colour (Fig. 2a). Where routes branched off from the trunk, they retained the letter and colour of the trunk, but acquired a numeric suffix. (Figure 2b shows the trunk-and-branch structure for the E train (red), formerly known as the IRT Seventh Avenue.) Stations would be uniquely identified by the route label, route colour, and the station name in signs displayed in stations and carriages (Fig. 3). That system was rejected, but we can see what might have been in Roberts' (2012) reconstruction. Although the TA made Salomon keep the three-colour principle, he did choose his own colours: key black (IRT), spot green (BMT), and spot red (IND). They paid $3000 for his map, but sought neither his involvement in managing the map, nor his signage, nomenclature, or colour coding. Salomon's proposed trunk-coloured map would have been clearer than the tricolour map, but the cost of changing the signage to match his trunk-and-branch nomenclature would have been prohibitive.



| Present name | New code letter and color | |
|---|---|---|
| IRT Lexington Avenue | B | 🟦 |
| BMT Broadway | C | 🟦 |
| IND Sixth Avenue | D | 🟫 |
| IRT Seventh Avenue | E | 🟥 |
| IND Eighth Avenue | F | 🟩 |

**Fig. 2.** (a) Trunks (b) branch scheme in Salomon's report (Color figure online)

**Fig. 3.** Station identifiers in Salomon's report

**1967: The D'Adamo Map.** A year after its birth, the TA announced a massive programme of infrastructure works to ease the major bottlenecks in the subway network (Ingalls 1954). One of these works was a two-mile tunnel under Chrystie Street, which was contracted nine years later (Anon 1963). That tunnel allowed the inter-working of trains on the former BMT and IND networks, which undermined the principle of a three-colour map that had been the common convention since the early 1930s.

By 1964, it was believed that completion was imminent and in late summer the TA opened up a Subway Map Contest to seek from the general public ideas on how best to revise the map, which they expected to need the following year.

Shaw (2011) suggested that it was not the Chrystie Street connection that triggered the Subway Map Contest, but the World's Fair, which New York City hosted from May 1964 to September 1965, and which led to a surge in the use of public transportation. In fact the TA had already put in place a comprehensive wayfinding programme by April (Anon 1964) including the 'blue streak' on the Salomon map, new route numbers on all buses, and new bus-stop signs (Perlmutter 1964), by the time they first mentioned a Subway Map Contest (Perlmutter 1964). The first year of the Fair closed in September when the Contest was ending, and they expected the new map to be out in autumn 1965. So the Map Contest can hardly have been aimed at the World's Fair.

Although the contest was intended to accommodate the inter-working of the BMT and IND networks, the materials sent to applicants made no mention of the Chrystie Street connection but included a copy of the 1964 pocket map as a reference. In May 1966, when the new map was quite advanced, Harold McLaughlin presented a paper on it at the annual meeting of the American Transit Association, but the TA immediately withdrew it and confiscated every copy they could find (Anon 1966), and it was omitted from the archives of the ATA. The TA remained reticent about the changes until very late, at which point it caused a lot of dissent, including attempted legal action to stop the opening of the Chrystie Street connection. It seems that the TA correctly expected a strong adverse reaction to the route changes that were concomitant with the opening of this new tunnel—which, as we shall see, had lasting ramifications.

In October 1964, the TA awarded $4000 to each of three winners (R. Raleigh D'Adamo, Harris Schechtman, and John & Mary Condon), but their maps were shelved and lost. (Fifty years later, Reka Komoli used a colour photograph to reconstruct D'Adamo's hand-drawn map as a vector file (Rhodes 2015).)

One winner, D'Adamo, submitted a report, explaining his principle of drawing each route in a distinct colour, and splicing together differently coloured routes running along a trunk. The TA hired Stanley Goldstein, a rocket scientist at Hofstra University, and passed D'Adamo's report to him. Goldstein submitted his report a year later: he and his students prepared four prototypes, and recommended #4, in which each route was drawn in a distinct colour (as proposed by D'Adamo) but routes running in parallel on a trunk were drawn side-by-side rather than spliced. Station stops were represented by squares (express) and circles (local), and transfers by proximity (the "no dot, no stop" rule). In #3 Goldstein reinvented Salomon's trunk-colour scheme: each trunk had a distinct colour, and each station was represented by a square in which was written the route codes of all the trains that stop there. Goldstein also took over D'Adamo's use of route identifiers in line-coloured rectangles at termini. In January 1966, Jerome Adler (Division Engineer in the TA Designs Division) decreed that the new map would combine Goldstein's prototypes #3 and #4. Each route would be drawn in a separate colour (as in #4) but each station was to be a box containing the route labels (as in #3). After a usability study in June (Barrington 1966), which yielded pink rectangles around transfers, the map passed to Diamond Packaging for editing and printing. There, Dante Calise selected the route colours and typeface, and the station maps were printed and installed for 26th November 1967, when the Chrystie Street opened (Fig. 4).

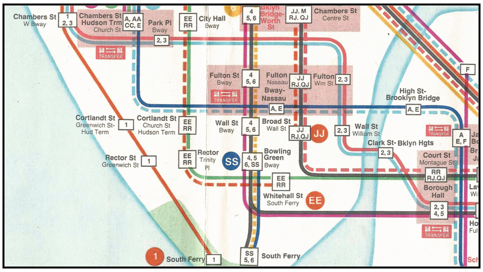


**Fig. 4.** Downtown excerpt from D'Adamo map, issued November 1967

**1967–1970: The Aftermath.** Although the new infrastructure eased the bottlenecks, the launch of the D'Adamo map was flawed. By announcing the changes just ten days before the opening (Perlmutter 1967), the TA left passengers no time to absorb the changes, or for the TA to absorb feedback. As only wall maps were printed on time, passengers had no pocket maps to study at home. By not updating the signage in subway cars and stations, they prevented passengers from relating the map to the platforms and services. The TA got many complaints, nominally about the new map but really prompted by the circumstances of its introduction. Also, the map itself was criticised: as a result of Adler's merging Goldstein's prototypes #3 and #4, the map was more fragmented and cluttered than necessary. D'Adamo himself sent in a critique of the new map, prompting TA to at least replace the pink boxes with clearer, station boxes.

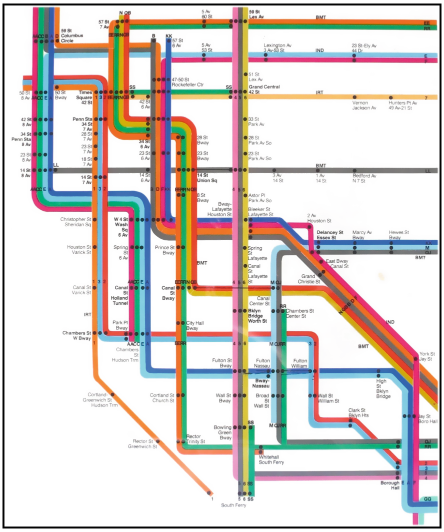


**Fig. 5.** 1970 comp: design director M. Vignelli, graphic designer J. Charysyn (Lloyd 2012)

**1972: The Vignelli Map.** A year after Chrystie Street, the TA was subsumed under a new state organization, the MTA, under the chair of William Ronan, and efforts soon commenced on a new subway map. What prompted the TA to seek a new subway map so soon after the three-year development of the D'Adamo map? Probably: (a) Bad press around the 1967 map might have motivated them to try again, this time with an outside firm rather than in-house. (b) As a new body, the MTA needed some early wins to build its brand in the public perception. (c) Unimark's signage project was concluding in 1970 with the release of the Graphic Standards Manual, and this created a natural opportunity to hire Unimark again to redesign the map as well. (d) Although the City of New York had only indirect influence over the TA, they were very critical of information delivery. For example, the City's Transportation Commissioner, wrote in the autumn of 1968, "The history of the TA's efforts to straighten out their graphics and designations is pitiful. […] We could proceed […] by simply telling the TA and the MTA that their present system is cockeyed and should be revised." (Sidamon-Eristoff 1968). Massimo Vignelli, head of the Unimark New York office, was already in touch with the TA on the signage project with Bob Noorda. He was scathingly critical of the 1967 map, and initiated a project to create a new, modernist map. With Joan Charysyn as graphic designer under Vignelli's direction, a comp was prepared by the summer of 1970 (Fig. 5), and quickly approved, with a contract signed between the TA and Unimark on 31st July 1970. The TA paid Unimark $17,600 for the map, but after it was issued in August 1972, neither Unimark nor Vignelli had any further involvement in the map. All modifications were handled in-house. In 1974, the map was completely redrawn, moving more of the map content into the empty north-east corner, and changing the typeface. A total of seven editions were issued (detailed by Lloyd 2012). The map was honoured as a 'design classic' and as 'iconic', but had vociferous critics who desired a return to a topographic map.

**1979: The Tauranac Map.** Ronan, who had championed the Vignelli map, was replaced in April 1974 by David Yunich, a Macy's marketing executive (Burks 1974). He created the MTA Marketing Department, and hired his former Macy's colleague Fred Wilkinson, who in 1975 formed the Subway Map Committee to supplant the Vignelli map with one that would lure in more passengers: subway maps had become primarily a marketing tool. For its first year, the Subway Map Committee had no vision of what should replace the Vignelli map, and even mooted a return to a tricolor scheme. In 1976, John Tauranac took the Chair, with an agenda of creating a topographic map,

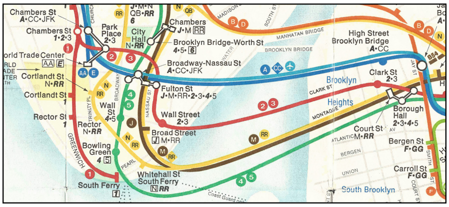


**Fig. 6.** Tauranac map, 1979

starting from that of the new guidebook (MTA 1976). With Tauranac as design director and Mike Hertz as graphic designer, and inputs from other members of the committee, by January 1978 a prototype map was publicly presented. As Tauranac knew, the map was flawed by using a single colour for all routes. In September, however, new funds became available and Tauranac was able to realise his vision of switching the subway from route colours to trunk colours, and hence deliver a topographically realistic map with trunk colour coding. This was issued on 25th June 1979 (Fig. 6). Basically the same concepts continue in the current MTA subway map: a trunk-coloured topographic map with the stopping routes listed alongside each station.

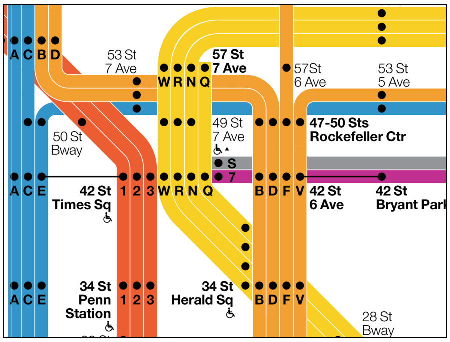


**Fig. 7.** Excerpt from MTA Weekender map

**2011: Waterhouse-Cifuentes Map.** In 2011, the MTA re-introduced a diagrammatic map in the style of Vignelli, designed by Yoshiki Waterhouse and Beatriz Cifuentes, to report temporary outages and re-routings because of engineering works. The intention was simply that the route-drawn diagram facilitated showing visually which individual routes were affected. This cannot be done visually in a trunk-drawn map such as Tauranac's, where outages must be listed as text. For example, in Fig. 7, if the N train had a weekend outage, then that route's line would be greyed out to show at a glance that service change. In the Tauranac map, this would require a textual note alongside each station where the N would normally stop. Originally existing only on the MTA Weekender web site (MTA 2017), the map is now routinely used in printed advisory notices that are displayed in stations.

## 3 Conclusion

The initial leap from geography to diagrams was driven chiefly by the TA's desire to cut costs and streamline map production, which fortuitously coincided with Salomon's long-standing desire for a clear London-style wayfinding system. The shift from the tri-colour company-based colour scheme to route colouring was driven by the need to keep the map legible after the BMT and IND merged. And the famous transformation into Vignelli's minimalist design was motivated by a corporate desire for rebranding after the creation of the MTA. Finally, the exit from the 'diagram decades' was instigated by Tauranac's vision of a 'didactic' map. Latterly, the Vignelli-style diagram was brought back because its separate route lines made it easier to show outages visually.

There is no grand narrative of the transitions of the diagrammatic subway map of New York City. Each change was made by individuals either to solve pragmatic problems or to express personal preferences. The simplistic notion that diagrammatic maps somehow do not suit New York is not supported by a close examination of the map's history, nor is the naïve notion that New York must inexorably follow an evolutionary trend from geographic maps to diagrams. Diagrammatic maps of the subway have specific advantages and disadvantages, and their comings and goings in NYC reflect this.

**Acknowledgement.** I am grateful to the following for permission to use their visuals: the Metropolitan Transportation Authority (Figs. 1, 5, 6 and 7) and the New York Transit Museum (Figs. 2, 3 and 4).

## References

Adler, J.: Memo to N.D. Brodkin. 18 January. NYTM Archives (1966)
Anon: Mayor Approves Contracts for Chrystie St, Subway, NY Times, 24 June 1963
Anon: Arrows in Subway to Aid Fair Visitors. NY Times, 8 April 1964
Anon: Subway Studying Color-Code Guide, NY Times, 30 May 1966
Barrington Inc.: A survey of prospective user reaction to the new subway map, 31 May. Report submitted to the NYCTA. Collection of Raleigh D'Adamo (1966)
BoT: Handwritten note detailing print runs and costs, 11 June. NYTM Archives (1952)
Burks, E.C: Yunich Now MTA's Chief Motorman. NY Times, 3 May 1974
Daly, W.J.: Memo to James H. Griffin, 3 July. NYTM Archives (1952)
Höhne, S.: How to make a map for the hades of names: NYC subway map wars of the 1970s. In: Cultural Histories of Sociabilities, Spaces & Mobilities. Taylor & Francis, pp. 83–98 (2015)
Ingalls, L.: Transit body asks $172,062,000 of city for capital improvements. NY Times, 10 July 1954
Lloyd, P.B.: Vignelli: Transit Maps. RIT Press, Rochester (2012)
MTA (Metropolitan Transportation Authority, New York, NY): The Weekender. http://web.mta.info/weekender.html. Accessed 14 Nov 2017
MTA: Seeing NY – The Official MTA Travel Guide. MTA, New York (1976)
Ovenden, M.: Transit Maps of the World. Penguin, New York (2015)
Perlmutter, E.: City Installing New Route Signs at Bus Stops, NY Times, 29 April 1964
Perlmutter, E.: Subway Changes to Speed Service. NY Times, 16 November 1967
Rawsthorn, A.: Hello World: Where Design Meets Life. Overlook Press, New York (2014)
Rhodes, M.: The lost NY subway map that may vastly improve modern ones, Wired Design, October 29 2015. https://www.wired.com/2015/10/the-lost-nyc-subway-map-that-may-vastly-improve-modern-ones/. Accessed 16 Feb 2018
Roberts, M.J.: Underground Maps Unravelled. Self-published, Wivenhoe (2012)
Roberts, M.J.: Henry Beck's London Underground Diagram and its Misuse and Use as a Cartographic Role Model. Accepted for publication in the Cartographic Journal (2018)
Salomon, F.: Email to the author, 27 April 2004
Salomon, F.: Interview with the author, Iowa City, 23 July 2015

Salomon, G.: Orientation. Typescript, May 1955. NYTM Archives (1955)
Salomon, G.: Out of the Labyrinth. Typescript. NYTM Archives (undated, inferred 1956a)
Salomon, G.: Letter to his father, 10 September. Translated by Edith Salomon (1956b)
Salomon, M.: Interview with the author, South Setauket, NY, 19 July 2006
Sidamon-Eristoff, C.: Memo to Mayor John Lindsey. 29 October. Municipal Archives (1968)
Shaw, P.: Helvetica and the New York City Subway System. MIT Press, Cambridge (2011)
Siegel, D.: Credit where Credit is due… or not. AIGA Design Observer, 19 May 2008
TA: Annual Report to 30 June, p. 22 (1958a)
TA: Annual Report to 30 June, p. 21 (1959)
TA: Letter from C.L. Patterson to Robert A. Abbott, September 9. NYTM Archives (1958b)
TA: Original Contract with Unimark International, 31 July. Vignelli Archives (1970)

# Sonifying Napoleon's March by Identifying Auditory Correlates of the Graphic-Linguistic Distinction

Peter W. Coppin[1](✉) and Richard C. Windeyer[2]

[1] OCAD University, Toronto, ON, Canada
`pcoppin@faculty.ocadu.ca`
[2] University of Toronto,Toronto, ON, Canada
`r.windeyer@mail.utoronto.ca`

**Abstract.** Identifying auditory correlates of the graphic-linguistic distinction informs our design of an auditory display based on Charles Minard's depiction of Napoleon's Russia campaign – the gold standard for visual (graphic) information design and therefore a grand challenge for auditory display design. We identify viable alternatives to the text-only translations currently employed in making graphics accessible to blind and/or low-vision individuals by introducing sounds bearing strong ecological resemblances to Minard's depictions. Our integration of theoretical work about classic distinctions with common properties across diagrammatic and auditory display communities reveals practical opportunities for designing inclusive and accessible graphics.

## 1 Translation from Graphics to Text for Accessibility

Charles Minard's *Figurative Map of the Successive Losses in Men of the French Army in the Russian Campaign 1812–1813* [19] (Fig. 1) is remarkable for representing six types of data in two dimensions [26]. Two "flow lines" represented the troops' journey – an upper tan-colored line traces the army's invasion of Russia, while a black lower line traces their retreat from Moscow back to Poland. The black "retreat" line is also linked to a temperature scale with dates and temperature readings (numbers). The positions of each point along the flow lines convey troop position (latitude and longitude), troop direction, and troop distance. Troop quantities are represented both by written numbers, at various points along the flow lines, and by the thickness of the flow lines themselves, with each millimeter of thickness representing 10,000 troops.

While a sighted individual can perceive the graphical content as an efficient integration of words, numbers and undulating graphical shapes, a blind or low-vision individual would typically access a sonic translation of it, consisting entirely of temporally sequenced streams of words (Fig. 2c).

Many blind and low-vision individuals depend on approaches like the Web Content Accessibility Guidelines (WCAG) to translate graphics into text descriptions which are then accessed aurally (through speech synthesis) via screen

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 228–235, 2018.
https://doi.org/10.1007/978-3-319-91376-6_23

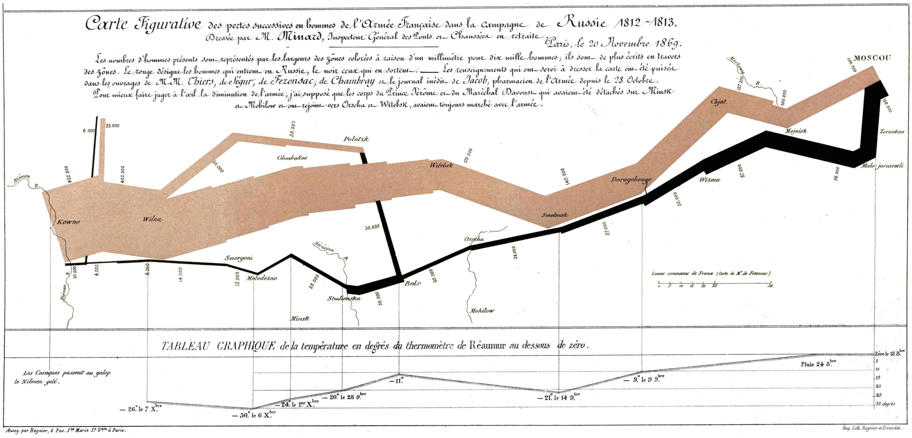
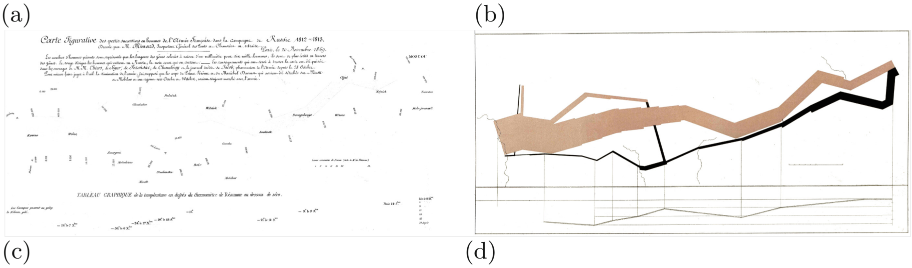


**Fig. 1.** Charles Minard's *Figurative Map of the Successive Losses in Men of the French Army in the Russian Campaign 1812–1813* (1869). Public domain.



(c)

River, and 100,000 men remained. They continued advancing towards the city of Moscow towards the northeast. At Moscow 100,000 men remained. They started retreating southwards towards the city of Malo-jarosewli, via the city of Tarantino. On October 18 they reached Malo-jarosewli and the temperature was zero degrees Réaumur. The remaining

(d)

advancing towards
towards the northeast.
retreating
southwards towards
On

**Fig. 2.** Deconstruction of Minard's map. The original map (Fig. 1) comprises parts conveyed via text (a) and parts conveyed via shapes (b), but in a text description (c) parts originally conveyed via shapes are also conveyed via text (d). Adapted from Charles Minard's *Figurative Map of the Successive Losses in Men of the French Army in the Russian Campaign 1812–1813* (1869). Public domain.

readers. Such text descriptions are essentially interpretations meant to convey the author's intended meaning of the infographic.

Certainly, the text labels conveying discrete troop quantities, isolated temperature readings and various landmarks (names of rivers, towns, battle sites) in [19] (Fig. 2a) lend themselves to text-only descriptions. However, the precise spatial relations between these elements across the 2D map – as well as their moment-by-moment changes – are conveyed graphically (Fig. 2b). Moreover, an important correlation between the changing temperature readings and troop quantities is also conveyed graphically.

This fundamental difference can be understood as the "graphic-linguistic distinction" [24] (G-L distinction). One might predict the parts originally conveyed via shapes (Fig. 2b) to be "lost in translation." Minimizing what is lost in translation, our *target design problem* requires identifying auditory correlates for the graphic-linguistic distinction. This is our *target theoretical problem.*

Although touch (cf. [13]) might be a more natural fit than sound, if the state of the art is primarily aural, a sound-based solution is probably more immediately applicable and aligns better with our target design problem. While screen readers can deliver text descriptions tactilely as braille, refreshable braille displays are expensive compared to audio hardware.[1] As well, [14] have shown that conveying information tactilely via touch screens also suffers from challenges related to haptic perception. Touch screens are also not as ubiquitous and inexpensive as audio hardware.

## 2 Identifying Auditory Correlates of the G-L Distinction

In [5], we reviewed four of the seven distinctions identified by Shimojima [24] and predicted what would be required for translation into the sonic domain, namely 2D versus Sequential, Relation Symbols versus Object Symbols, Analog Versus Digital and Intrinsic versus Extrinsic Constraints.

### 2.1 2D vs. Sequential, Analog vs. Digital and Intrinsic vs. Extrinsic

Larkin and Simon [16] distinguish between diagrammatic ("a data structure in which information is indexed by two-dimensional location") and sentential representations ("a data structure in which elements appear in a single sequence") (p. 68): in Minard's graphic, by definition, the text (Fig. 2c) is sentential, due to its arrangement as a linear sequence of marks, whereas the marks that are indexed to the 2D plane are diagrammatic. The spatial relations among the labeled marks enable visual perception of the contour of the line or the relative positions of marks scattered across the 2D surface, conveying values and trends that are not conveyed via labels (cf. [1]).

The sentential properties of Minard's graphic are easily conveyed via text-to-speech. However, for **sonic diagrammatics**, in [5] we predicted that the formation of a 2D space in the sonic domain requires two properties of sound that can be independently manipulated and perceived.

---

[1] The Orbit Reader, "the first ever affordable refreshable braille reader that is portable," costs $499 Canadian. As of December 2017 the device is still only available as a pre-order [4].

**Auditory Display Design.** Just as graphic designers arrange combinations of visual variables (lines, shapes, colors, textures) across a 2D surface to convey information, auditory display designers employ sonic variables including pitch, loudness, spectral brightness (or timbre), tempo, duration and spatial position within a stereo (or surround-sound) listening environment [15,17,21,25].

Thus, our design solution translates the 2D visual plane of Minard's graphic to a horizontal, auditory "ground plane" so as to recruit a fundamental affordance of human hearing – the ability to localize and isolate sonic events within a 360° radius. Our selected auditory icons (sonic correlates on the graphic side and detailed next), earcons (correlates on the linguistic side) and text-to-speech label translations (conveying troop quantities, geographic landmarks and temperature readings) are then mapped to the horizontal auditory plane by translating the vertical and horizontal coordinates of their corresponding markings on the visual map to the corresponding azimuth and distance coordinates of the auditory ground plane. From a fixed listening position (the auditory equivalent to the centre point of the infographic), the marching footsteps of the advancing army move from left to right towards Moscow in front of the listener, while the marching footsteps of retreating army move from right to left, behind the listener. Simultaneously, text-to-speech descriptions denoting discrete troop quantities, geographic landmarks, and temperature readings are also spatially mapped to their corresponding coordinates on the infographic.

The foregoing also aligns with the distinction between **Intrinsic and Extrinsic Constraints** [24]: representations that obey "inherent constraints" to be graphical (p. 332). Drawing from [1], Shimojima reinforces the notion that "diagrams are physical situations" (p. 22) that adhere to their own intrinsic set of constraints. When the diagram's constraints are appropriately matched to the constraints of the situation described, an appropriate representational has been chosen.

As noted above, we recruited a fundamental affordance of human hearing – the ability to localize and isolate sonic events within a 360° radius. Via our sonic translation, the horizontal constraints of a stereo or binaurally-encoded audio environment map appropriately to the constraints of the infographic's horizontal dimensions, thus utilizing intrinsic constraints. Likewise, the constraints of pitch/frequency are appropriate for representing vertically-spaced relations between markings within various categories, such as the temperature readings.

**Correlates of Analog Versus Digital.** The Analog Versus Digital distinction (most commonly associated with [10]) describes analog systems as continuous and capable of changing the meaning of a representation by placing new elements between any two other elements ad infinitum. By contrast, digital systems are discontinuous and differentiated throughout. Under this distinction, pictures are considered analog and more replete; diagrams are also analog though less replete; and linguistic systems are partially digital. As linguistic-symbolic properties bear little or no resemblance to the item they represent, they are described as having a conventionalized relationship with the items they represent. In the context of auditory display design, this description most closely corresponds to the arbitrary or conventionalized relationships that earcons typically have with the data-derived items they represent.

In distinguishing the properties of auditory icons and earcons, [15] established a representational continuum with the "analogic" properties of auditory icons at one end and the "symbolic" properties of earcons at the other. [7,9] further defined this continuum by establishing a taxonomy comprised of the signal-referent relations described above.

Overall, this approach to auditory mapping aims to preserve Larkin and Simon's principle observation about the role of diagrammatic representations in explicitly preserving information about topographical and geometric relations among the components..." (p. 66). (Figure 2, upper right) and [ANALG DIG].

## 2.2 Relation Symbols and Object Symbols

Russell's distinction between *Relation Symbols and Object Symbols* [23] identifies "words which mean relations are not themselves relations, but just as substantial or insubstantial as other words. In this respect, a map is superior to language, since one location is to the west of another is represented by the fact that the corresponding place on the map is to the left of the other; that is to say, a relation is represented by a relation." (p. 90)

Minard's graphic includes relations among *troop locations*, *troop quantities*, and *temperatures*. **Relations among troop locations** are conveyed by relation symbols that consist of visually perceived relations among points on the x or y-axis of the graphic where the various points on the x or y-axis represent the relations among troop locations relative to space-time. In the visual version, **relations among troop quantities** relative to space-time are conveyed by visually perceiving relations among line thicknesses relative to distance. In the visual version, **relations among temperatures** over space-time are conveyed by visually perceiving relations among line elevations relative to locations on the y-axis.

**Auditory Display Design of Relation Symbols.** *Parameter Mapping Sonification* describes how selected dimensions of a dataset may be represented sonically as perceivable changes to one or more sonic variables [3,11]. Usability testing reveals the relative strengths and weaknesses of each sonic variable in representing dimensions of data in different applications and design situations [2,8,20]. In representing temperature changes for example, pitch is considered more effective than tempo [27], whereas loudness is considered to be much less effective [22]. Effective combinations of sonic variables are referred to as earcons - "short, structured musical phrases that can be parameterized to communicate information in an Auditory Display" ([18], p. 339).

By contrast, *parameterized auditory icons* [3] are digitally recorded or computationally generated sounds whose sonic characteristics, behaviors and variabilities bear a direct or indirect ecological or metaphorical resemblance to a target referent – real-world sounds [12] and one or more dimensions of data to be represented. *Direct signal-referent relations* denote sounds possessing a direct resemblance to the target referent. *Indirect signal-referent relations* provide a "surrogate sound" for the target referent. Indirect relations are further subdivided into two subcategories: *Indirect-ecological relations* denote surrogate sounds and target referents that actually coexist in the world. *Indirect-metaphorical relations* denote surrogate sounds and target referents that relate to each other through similarities in audio-visual appearance or functions.

By way of example, our sonic translation employs two instances of parameterized auditory icons. The first possesses *direct-ecological signal-referent relations* – a parameterized recording of footsteps marching on a variety of seasonally related ground surfaces. Larger troop quantities are represented by layering multiple instances of the auditory icon to create a sonic representation of Napoleon's army. As troop quantities diminish and seasonal temperatures drop, these layers are reduced, marching tempo slows, and audible changes to the surfaces underfoot occur (i.e., gravel, water, ice, snow). The second auditory icon possesses *indirect-metaphorical signal-referent relations* – a computationally generated sound of a container filled with small objects being shaken. The number of objects heard rattling around inside the container is mapped to changes in troop quantity. Importantly, our mapping of the marching footsteps auditory icon attempts to address the challenge of preserving information conveyed via spatial relations between graphical shapes rather than labels (cf. [1]).

## 3 Conclusion

Our design prototype explores the 2D versus sequential distinction through the use of binaural (surround-sound) encoding methods to assign precise spatial coordinates to each auditory icon, earcon, or text-to-speech label. In doing so, the spatial relations between markings on the infographic – and their communication of spatial, geometric, or topological information – is preserved. Our spatial positioning of auditory icons offers an additional affordance – characterized by the relation symbols and object symbols distinction – wherein continuously changing numerical values (troop quantities, locations) may be mapped to perceptual dimensions. This suggests that analog and spatial properties of sound could be recruited in mapping numerical values to perceptual dimensions. This highly spatialized approach to sonic translation is further strengthened by matching the intrinsic constraints (versus extrinsic constraints) of the infographic's physical situation with the intrinsic constraints of spatial auditory perception capabilities in the majority of potential users. Finally the our selection of parameterized auditory icons (marching footsteps), earcons (descending pure tone representing changing temperatures), and text-to-speech label translations illustrates distinctions between analog and digital systems.

**Acknowledgments.** This research was supported by funding from the Social Science and Humanities Research Council of Canada and the Natural Sciences and Engineering

Research Council of Canada, and The Centre for Innovation in Data-Driven Design. We would like to thank research assistants Ambrose Li, Michael Carnevale, Damon Pfaff, Brandon Biggs and Lena Yusim as well as Marta Wnuczko, Dr. David Steinman and Daniel E. MacDonald.

## References

1. Barwise, J., Etchemendy, J.: Visual information and valid reasoning. In: Zimmerman, W. (ed.) Visualization in Mathematics, pp. 8–23. Mathematical Association of America, Washington (1990)
2. Bonebright, T.L., Nees, M.A., Connerley, T.T., McCain, G.R.: Testing the effectiveness of sonified graphs for education: a programmatic research project. In: Proceedings of International Conference on Auditory Display (ICAD2001), pp. 62–66, Espoo, Finland (2001)
3. Brazil, E., Fernström, M.: Auditory icons. In: Hermann, T., Hunt, A., Neuhoff, J.G. (eds.) The Sonification Handbook, pp. 325–338. Logos, Berlin (2011)
4. CNIB: Deposit for Orbit Braille Reader 20. http://shop.cnib.ca/ProductDetail/tec9999999999_deposit-for-orbit-braille-reader-20
5. Coppin, P.W., Li, A., Carnevale, M.: Iconic properties are lost when translating visual graphics to text for accessibility. In: Zlatev, J., Konderak, P., Sonesson, G. (eds.) Meaning, Mind, and Communication: Explorations in Cognitive Semiotics, pp. 279–295. Peter Lang, Frankfurt am Main (2016)
6. Coppin, P.W., Windeyer, R.C., MacDonald, D.E., Steinman, D.A.: Progress toward sonifying Napoleon's March and fluid flow simulations through binaural horizons. In: International Conference on Auditory Display (2017). https://doi.org/10.21785/icad2017.042
7. Familant, M.E., Detweiler, M.C.: Iconic reference: evolving perspectives and an organizing framework. Int. J. Man-Mach. Stud. **39**, 705–728 (1993)
8. Flowers, J.H., Hauer, T.A.: Musical versus visual graphs: cross-modal equivalence in perception of time series data. Hum. Factors **37**, 553–569 (1995)
9. Gaver, W.W.: Using and creating auditory icons. In: Kramer, G. (ed.) Auditory Display: Sonification, Audification and Auditory Interfaces, pp. 417–446. Addison-Wesley, Reading (1994)
10. Goodman, N.: Languages of Art: An Approach to a Theory of Symbols. Bobbs-Merrill Company, Indianapolis (1968)
11. Grond, F., Berger, J.: Parameter mapping sonification. In: Hermann, T., Hunt, A., Neuhoff, J.G. (eds.) The Sonification Handbook, pp. 363–397. Logos Publishing House, Berlin (2011). http://sonification.de/handbook
12. Keller, P., Stevens, C.: Meaning from environmental sounds: types of signal-referent relations and their effect on recognizing auditory icons. J. Exp. Psychol. Appl. **10**, 3–12 (2004). https://doi.org/10.1037/1076-898X.10.1.3
13. Kennedy, J.M.: Drawing and the Blind. Yale University Press, New Haven (1993)
14. Klatzky, R.L., Giudice, N.A., Bennett, C.R., Loomis, J.M.: Touch-screen technology for the dynamic display of 2D spatial information without vision: promise and progress. Multisensory Res. **27**, 359–378 (2014). https://doi.org/10.1163/22134808-00002447
15. Kramer, G.: An introduction to auditory display. In: Kramer, G. (ed.) Auditory Display: Sonification, Audification and Auditory Interfaces, pp. 1–78. Addison-Wesley, Reading (1994)

16. Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. Cogn. Sci. **11**, 65–99 (1987). https://doi.org/10.1111/j.1551-6708.1987.tb00863.x
17. Levitin, D.J.: Memory for musical attributes. In: Cook, P. (ed.) Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics, pp. 209–227. MIT Press, Cambridge, MA (1999)
18. McGookin, D., Brewster, S.: Earcons. In: Hermann, T., Hunt, A., Neuhoff, J.G. (eds.) The Sonification Handbook, pp. 339–361. Logos, Berlin (2011). http://sonification.de/handbook
19. Minard, C.J.: Carte figurative des pertes en hommes de l'Armée Française dans la campagne de Russie 1812–1813. In: Tableaux graphiques et cartes figuratives de Mr. Minard, p. 28. Regnier et Dourdet, Paris (1869). https://patrimoine.enpc.fr/document/ENPC01_Fol_10975?image=54
20. Neuhoff, J.G., Heller, L.M.: One small step: sound sources and events as the basis for auditory graphs. In: Proceedings of 11th Meeting of the International Conference on Auditory Display, Limerick, Ireland (2005)
21. Neuhoff, J.G.: Perception, cognition and action in auditory display. In: Hermann, T., Hunt, A., Neuhoff, J.G. (eds.) The Sonification Handbook, pp. 63–85. Logos Publishing House, Berlin (2011). http://sonification.de/handbook
22. Neuhoff, J.G., Kramer, G., Wayand, J.: Pitch and loudness interact in auditory displays: can the data get lost in the map? J. Exp. Psychol. Appl. **8**, 17–25 (2002)
23. Russell, B.: Vagueness. Australas. J. Psychol. Philos. **1**, 84–92 (1923)
24. Shimojima, A.: The graphic-linguistic distinction: exploring alternatives. Artif. Intell. Rev. **13**, 313–335 (1999). https://doi.org/10.1023/A:1006544013197
25. Sodnik, J., Tomažič, S.: Spatial Auditory Human-Computer Interfaces. SCS. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22111-3
26. Tufte, E.R.: The Visual Display of Quantitative Information. Graphics Press, Cheshire (2001)
27. Walker, B.N.: Consistency of magnitude estimations with conceptual data dimensions used for sonification. Appl. Cogn. Psychol. **21**, 579–599 (2007)
28. Coppin, P.W.: What is lost in translation from visual graphics to text for accessibility. In: Proceedings of EuroAsianPacific Joint Conference on Cognitive Science, vol. 26, p. 27
29. Windeyer, R.C., Coppin, P.W.: Sonification Prototype 2.0: a binaural translation of Charles Minard's infographic "Figurative Map of the Successive Losses in Men of the French Army in the Russian campaign 1812–1813" (1869) (2018). https://vimeo.com/217423986

# Towards a Typology of Diagrams in Linguistics

Hans Smessaert[1(✉)] and Lorenz Demey[2]

[1] Department of Linguistics, KU Leuven, Leuven, Belgium
Hans.Smessaert@arts.kuleuven.be
[2] Center for Logic and Analytic Philosophy, KU Leuven; FWO, Leuven, Belgium
Lorenz.Demey@hiw.kuleuven.be

**Abstract.** The aim of this paper is to lay out the foundations of a typology of diagrams in linguistics. We draw a distinction between linguistic parameters — concerning *what* information is being represented — and diagrammatic parameters — concerning *how* it is represented. The six binary linguistic parameters of the typology are: (i) mono- versus multilingual, (ii) static versus dynamic, (iii) mono- versus multimodular, (iv) object-level versus meta-level, (v) qualitative versus quantitative, and (vi) mono- versus interdisciplinary. The two diagrammatic parameters are (i) iconic/concrete versus symbolic/abstract representation and (ii) static versus dynamic representation. We briefly illustrate how different types of linguistic diagrams can be analysed in terms of the interaction between the linguistic and the diagrammatic parameters.

**Keywords:** Typology · Linguistics · Modularity
Object vs metalevel · Iconic representation · Symbolic representation

## 1 Diagrams and Linguistics

In order to put the present paper in a somewhat broader context, let us start off by briefly considering three different perspectives which the Diagrams research community has taken upon the relationship between diagrams on the one hand and language or linguistics on the other hand. The first perspective — namely *Diagrams versus Linguistics* — adopts a primarily 'negative' relation of contrast between the two. The second and third perspectives both establish 'positive' relations, but they differ from one another in terms of their directionality, namely *Linguistics for Diagrams* as opposed to *Diagrams for Linguistics*.

**Diagrams versus Linguistics.** At least since the seminal paper of Larkin and Simon [9], research into reasoning systems – at the interface of Logic, Cognitive Science and Artificial Intelligence — has been concerned with similarities and differences between reasoning based on diagrammatic or visual information and reasoning based on sentential, propositional or linguistic information [5,16,18,19]. Also outside the context of reasoning research, diagrammatic information

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 236–244, 2018.
https://doi.org/10.1007/978-3-319-91376-6_24

is often distinguished from, and opposed to, linguistic information, e.g. when diagrams are said to occupy an intermediate position in between ‘description’ and ‘depiction’ [11], or when the category of ‘non-picture visuals’ is further divided into ‘linguistic forms’ — such as text, tables or notations — and ‘non-picture graphical forms’ — such as diagrams, graphs, charts and maps [4].

**Linguistics for Diagrams.** One way of establishing a more positive relation between a diagrammatic and a linguistic perspective involves using concepts from the field of linguistics to study or explain properties of diagrams. The key idea of the research field on Visual Languages and Computation is precisely that graphical representations resemble (natural or formal) languages in having a vocabulary and a grammar. As Mackinlay put it, “graphical presentations are actually sentences of graphical languages that have precise syntactic and semantic definitions” [10]. This concept of the ‘grammar’ or ‘language’ of graphics and visual design is worked out in great detail by Kress and van Leeuwen [8] as well as Engelhardt [6]. Interestingly, the first two perspectives on the relationship between diagrams and linguistics are integrated in Howse et al. [7], where the linguistic concept of a type-token distinction is taken to play a bigger role in diagrammatic systems than in linguistic systems, thus arguing for a more fine-grained syntax for diagrammatic representations.

**Diagrams for Linguistics.** An alternative way of connecting diagrams and linguistics proceeds by taking concepts from the field of diagrams research in order to study the visual representations used in linguistics to describe the properties of natural language expressions. Judging from the contributions to the proceedings of the nine Diagrams Conferences (2000–2016), this third perspective has received surprisingly little attention. The ones that do occur, however, testify of the fact that diagrams show up in various areas of linguistic research, ranging from phonological features [14], over tree representations for syntactic structures [3], to scales and sets for the semantic representation of tenses [2] and quantifiers [15].

**Aim of the paper.** The aim of the present paper is precisely to contribute to this third perspective, by laying out the foundations of a typology of diagrams in linguistics. To the best of our knowledge, such a typology has not been proposed so far. We draw a distinction between *linguistic* parameters in Sect. 2 — concerning *what* information is being represented — and *diagrammatic* parameters in Sect. 3 — concerning *how* that information is represented. In Sect. 4 we briefly illustrate how different types of linguistic diagrams can be analysed in terms of the interaction between linguistic and diagrammatic parameters. In future work, the validity of the proposed parameters will be empirically tested by means of large-scale corpus research. In the long run, the resulting typology is hoped to contribute to the emerging field of philosophy of linguistics, and hence to the broader area of philosophy of science, in which the heuristic and didactic value of visualisation techniques is a well-established research topic [17].

## 2 Linguistic Parameters

As to the question of *what* information is being represented, the first four binary parameters of the typology receive an intrinsically linguistic characterisation, namely: (i) mono- versus multilingual, (ii) static versus dynamic, (iii) mono- versus multimodular, and (iv) object-level versus meta-level information. The remaining two binary parameters — (v) qualitative versus quantitative, and (vi) mono- versus interdisciplinary information — are of a more general nature.

**Monolingual versus multilingual information.** The first parameter distinguishes between the — MONOLINGUAL — study of 'language' and the— MULTILINGUAL — study of 'languages'. The study of language refers to expressions in a particular natural language — such as the simple English main clause *The cat is sitting on the mat* — on different levels of complexity. This complexity is traditionally related to the 'size' of the units under scrutiny, ranging from very small to very big, in particular from sounds over words and clauses to discourse. The study of languages, by contrast, is concerned with family relationships between (groups of) natural languages — such as the Germanic versus the Celtic language families. This field of study — often called 'linguistic typology' — crucially involves the dimensions of space and time, since it aims to chart the geographical distribution of language families as well as their chronology, i.e. their genetic resemblance and descendance. Both the monolingual analysis of natural language expressions and the multilingual classification and comparison of language families very often make use of visual representations of various kinds.

**Static versus dynamic information.** The second parameter — which concerns the opposition between STATIC and DYNAMIC information — can first of all straightforwardly be connected to the linguistic contrast between synchrony and diachrony. The *synchronic* perspective considers the contemporary situation both 'internally', for any given individual language, and 'externally', for language families as a whole. The *diachronic* perspective, by contrast, investigates the historical changes and evolutions, again both language-internally and on the level of entire language families. It is important to stress, however, that the distinction between static and dynamic information is not restricted to the synchronic versus diachronic perspectives in linguistics. In general, static information concerns a stable situation or state of an object or concept, whereas dynamic information concerns processes, i.e. temporal or structural changes in the object or concept. For instance, in theoretical frameworks which assume that certain components of a natural language expression are moved to different positions in the structure.

**Monomodular versus multimodular information.** As already hinted at above, natural language expressions can be analysed on different levels, depending on the size of the units or components under investigation. Standardly, the following six linguistic modules are distinguished:

| | |
|---|---|
| *phonetics* | the articulation, acoustics and perception of speech sounds |
| *phonology* | abstract sound segments, syllables and prosody |
| *morphology* | the internal structure of words (derivation, compounding) |
| *syntax* | sentence structure and word-order |
| *semantics* | the meaning of words (lexical) and sentences (propositional) |
| *pragmatics* | speech acts, interaction in conversation, discourse structure |

According to the third parameter, MONOMODULAR information restricts the focus to properties, relations or concepts within one of the six modules above, whereas MULTIMODULAR information is at issue as soon as properties, relations or concepts from at least two modules are shown to interact in the representations.

**Object-level versus meta-level information.** As a fourth parameter, we propose a binary opposition between *object-level* information and *meta-level* information. In linguistics, the key 'objects' of investigation are natural language expressions: sounds, syllables, words, constituents, clauses, conversations and so on. Visual representations that explicitly contain (components of) such natural language expressions and their properties will be called OBJECT-LEVEL diagrams. Quite often, however, visual representations are concerned with properties of, or relations between, the linguistic concepts or the applied terminology, irrespective of any concrete natural language expression. Such representations will be called META-LEVEL diagrams.

**Qualitative versus quantitative information.** With the fifth parameter — which relates to the opposition between *qualitative* and *quantitative* information — we reach a more general, no longer intrinsically linguistic level. On the one hand, the analysis of properties of natural language expressions (object-level) or relations between linguistic concepts (meta-level) often yields QUALITATIVE or non-numerical data/information, concerning — for instance — linear ordering relations or hierarchical structures. On the other hand, linguistic analyses very often also generate QUANTITATIVE information, in the form of numerical values for certain parameters or attributes (or any other statistical properties). The latter receive visual representations such as tables or charts (bar charts, line charts, pie charts and so on) which are omnipresent in the scientific literature, but not specifically linguistic in nature [13, p. 61].

**Monodisciplinary versus interdisciplinary information.** As is the case in so many scientific areas, the field of linguistics interacts with a whole range of neighbouring disciplines, such as sociology, psychology, neuroscience, computer science (among many others), thus giving rise to the corresponding interdisciplinary fields of sociolinguistics, psycholinguistics, neurolinguistics, and computational linguistics. The sixth and final linguistic parameter therefore distinguishes between MONODISCIPLINARY information on the one hand — i.e. 'purely linguistic' information restricted to the core modules (or levels) for the analysis of linguistic expressions — and INTERDISCIPLINARY information on the other hand. After all, since those many interdisciplinary fields not only borrow concepts and methods from their non-linguistic source disciplines, but very often also adopt the corresponding visualisation strategies, we will need to take visualisations of 'mixed origin' into account as well.

## 3 Diagrammatic Parameters

Following the characterisation by Purchase [13, p. 59], a DIAGRAM is taken to be "a composite set of marks or visual elements — including lines, geometric shapes and individual words — on a two-dimensional plane, that — when taken together — represent a concept or object in the mind of the viewer. Diagrams are meant to depict appearance, structure, or workings of something, and are usually employed to support viewers' tasks, such as learning, designing, communicating, or simply understanding the concept depicted."

As a number of overview papers have demonstrated [1,13], it is notoriously difficult to provide an adequate set of visual principles for setting up a typology of diagrams. Nevertheless, as was the case in the previous section, a number of binary oppositions turn out to play a crucial role. In order to answer the question of *how* the linguistic information is being represented visually, two diagrammatic parameters are taken to underlie the typology. The main diagrammatic parameter is the classical, semiotic contrast between iconic/concrete and symbolic/abstract representations. On a secondary level, we draw a further distinction between static and dynamic representations.

**Iconic versus symbolic representation.** The central diagrammatic parameter to underlie the envisaged typology, is taken from the field of semiotics (the science of signs), namely the distinction between two major classes of signs — ICONS and SYMBOLS — which basically corresponds to the opposition between concrete and abstract representations [13, p. 59]. A diagram is called ICONIC or CONCRETE if there is a direct perceptual relationship of similarity or resemblance between the *sign* — i.e. the diagram or representation — and the *referent* — i.e. the object being represented. A diagram is called SYMBOLIC or ABSTRACT if the relationship between the sign and the referent is purely arbitrary, based on sets of conventions within a given community.

ICONIC diagrams depict their objects in a form similar to their physical attributes or depict physical positional relationships between objects [13, p. 60]. Typical examples are anatomic illustrations, maps, or seating arrangements. SYMBOLIC diagrams, by contrast, have no perceptual relationship to the concepts that they represent. Three broad categories can be distinguished [13, p. 60]: (i) *graphs* use geometric shapes to represent objects, and lines to depict relationships between objects, (ii) *set diagrams* use overlapping geometric shapes to depict set membership, and (iii) *charts* — as mentioned above — present numerical or quantitative information[1]. Note that specific diagrams may be of a *composite* type in that they combine notational properties of different (abstract and/or concrete) diagram subtypes [13, p. 61].

---

[1] In the field of Information Graphics or Data Visualisation, the different ways in which information can be structured have been captured under the acronym LATCH (= Location Alphabet Time Category Hierarchy), according to whether the elements are organised spatially, organised alphabetically, organised against a time line, divided into classes or ranked in order of priority [21]. The Location dimension typically yields concrete diagrams, whereas the others standardly yield abstract diagrams.

Within the symbolic category of *graphs*, the subtype of *trees* — think of the classical genealogical or family tree — deserves special mention here. The notion of hierarchical structure obviously plays a crucial role in many scientific disciplines [12,20]. Also in the field of linguistics, trees pop up all over the place, first of all in the genealogical sense, as representations of the internal structure of language families. Secondly, also on the level of language — i.e. the properties of concrete natural language expressions — the idea of hierarchical organisation is absolutely essential, with varying sizes for the units of analysis, depending on the particular linguistic module under consideration.

**Static versus dynamic representation.** Independently of the above opposition between iconic and symbolic diagrams, we will also distinguish between STATIC diagrams and DYNAMIC diagrams [13, p. 62]. A DYNAMIC diagram represents a succession or sequence of states or processes, either as a series of individual diagrams[2], or by using graphical elements such as arrows. By contrast, a STATIC diagram contains no graphical elements representing a succession or sequence. Note that it is possible for a static diagram to represent dynamic information, as long as it does not contain any dedicated graphical elements for this purpose (such as arrows). However, such a mismatch between the diagram and the information that it represents typically has a negative effect on the quality of the visual representation.

## 4 Illustrating and Applying the Parameters

In this section we briefly illustrate how different types of linguistic diagrams can be analysed in terms of the interaction between the linguistic parameters (LP) from Sect. 2 and the diagrammatic parameters (DP) from Sect. 3. Figure 1(a) iconically represents the tongue movement in the vowel space of the two dipthong sounds in Dutch words such as ***bruin*** (‘brown’) and ***koud*** (‘cold’), whereas the spectrogram (from acoustics/physics) in Fig. 1(b) provides quantitative information on the distribution of frequency over time (with intensity as gray scale), when pronouncing the Dutch expression *aja* (‘yes indeed’). Figure 1(c) represents the historical development of the Insular branch of the Celtic language family, and Fig. 1(d) represents the multimodular mismatches between the syntactic and morphological structures of the West Flemish subclause … *da-k eet-n* (‘…that I am eating’). In Table 1, the linguistic and diagrammatic parameters proposed above, are applied to the four linguistic diagrams from Fig. 1. Notice that the genealogical tree structure in Fig. 1(c) is a typical example of dynamic information (viz. the evolution of certain languages over time) being represented by means of a static diagram.

---

[2] Notice that such a series of diagrams can develop as an *animation* through time, or by *juxtaposition* in space.

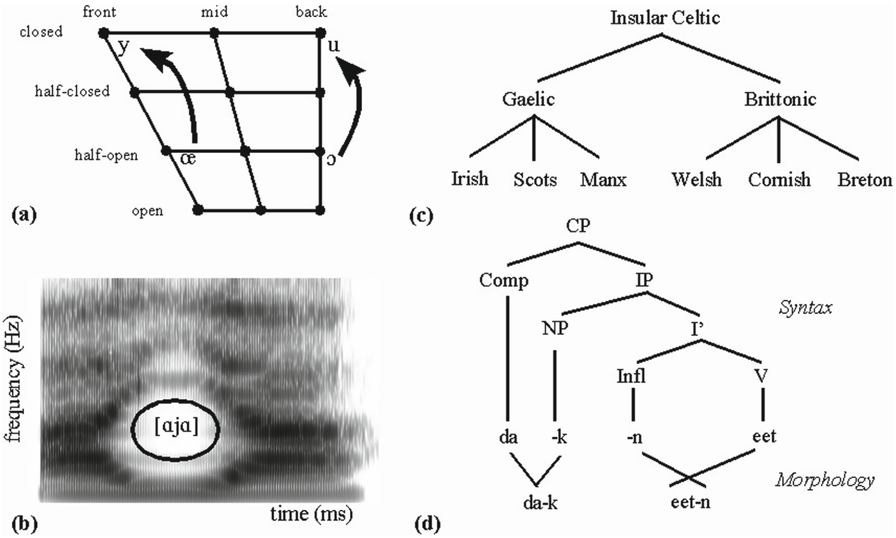


**Fig. 1.** Diagrams in Linguistics

**Table 1.** Linguistic and diagrammatic parameters applied to Fig. 1

| | [–] | [+] | (a) | (b) | (c) | (d) |
|---|---|---|---|---|---|---|
| LP | Monolingual | Multilingual | – | – | + | – |
| | Static | Dynamic | + | + | + | – |
| | Monomodular | Multimodular | – | – | – | + |
| | Object-level | Meta-level | – | – | + | – |
| | Qualitative | Quantitative | – | + | – | – |
| | Monodisciplinary | Interdisciplinary | – | + | – | – |
| DP | Iconic | Symbolic | – | + | + | + |
| | Static | Dynamic | + | + | – | – |

## 5 Conclusions and Future Work

In this paper we have laid out the foundations of a typology of diagrams in linguistics by looking at the interaction between six linguistic parameters and two diagrammatic parameters. In future work, the validity of the proposed parameters will be empirically tested by means of large-scale corpus research, based on a broad range of general linguistics journals and handbooks dedicated to the various linguistic modules. This will allow us to check whether the proposed parameters are sufficiently fine-grained to capture and classify all actually occurring diagrams, or whether further modifications are necessary. Ultimately, the resulting typology will help to clarify the heuristic and didactic value of visualisation techniques in linguistics, thus contributing to the emerging field of philosophy of linguistics.

## References

1. Blackwell, A., Engelhardt, Y.: A meta-taxonomy for diagram research. In: Anderson, M., Meyer, B., Olivier, P. (eds.) Diagrammatic Representation and Reasoning, pp. 47–64. Springer, London (2002). https://doi.org/10.1007/978-1-4471-0109-3_3
2. Bosveld-de Smet, L.: Diagrams in second or foreign language learning??! In: Barker-Plummer, D., Cox, R., Swoboda, N. (eds.) Diagrams 2006. LNCS (LNAI), vol. 4045, pp. 151–153. Springer, Heidelberg (2006). https://doi.org/10.1007/11783183_20
3. Bosveld-de Smet, L., de Vries, M.: Visualizing non-subordination and multidominance in tree diagrams: testing five syntax tree variants. In: Stapleton, G., Howse, J., Lee, J. (eds.) Diagrams 2008. LNCS (LNAI), vol. 5223, pp. 308–320. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87730-1_28
4. Cox, R.: Recognising, knowing and naming: can object picture processing models accommodate non-picture visuals? In: Dwyer, T., Purchase, H., Delaney, A. (eds.) Diagrams 2014. LNCS (LNAI), vol. 8578, pp. 161–175. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44043-8_19
5. Cox, R., Dale, R., Etchemendy, J., Barker-Plummer, D.: Graphical revelations: comparing students' translation errors in graphics and logic. In: Stapleton, G., Howse, J., Lee, J. (eds.) Diagrams 2008. LNCS (LNAI), vol. 5223, pp. 257–265. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87730-1_24
6. Engelhardt, Y.: The language of graphics: a framework for the analysis of syntax and meaning in maps, charts and diagrams. Ph.D. thesis, Universiteit van Amsterdam, Institute for Logic, Language and Computation (2002)
7. Howse, J., Molina, F., Shin, S.-J., Taylor, J.: On diagram tokens and types. In: Hegarty, M., Meyer, B., Narayanan, N.H. (eds.) Diagrams 2002. LNCS (LNAI), vol. 2317, pp. 146–160. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-46037-3_18
8. Kress, G.R., van Leeuwen, T.: Reading Images: The Grammar of Visual Design. Routledge, Abingdon (1996)
9. Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. Cogn. Sci. **11**(1), 65–100 (1987)
10. Mackinlay, J.D.: Automating the design of graphical presentations. ACM Trans. Graph. **5**(2), 110–141 (1986)
11. Norman, J.: Differentiating diagrams: a new approach. In: Anderson, M., Cheng, P., Haarslev, V. (eds.) Diagrams 2000. LNCS (LNAI), vol. 1889, pp. 105–116. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-44590-0_13
12. Novick, L.R.: The importance of both diagrammatic conventions and domain-specific knowledge for diagram literacy in science: the hierarchy as an illustrative case. In: Barker-Plummer, D., Cox, R., Swoboda, N. (eds.) Diagrams 2006. LNCS (LNAI), vol. 4045, pp. 1–11. Springer, Heidelberg (2006). https://doi.org/10.1007/11783183_1
13. Purchase, H.C.: Twelve years of diagrams research. J. Vis. Lang. Comput. **25**(2), 57–75 (2014)
14. Salting, D.: Feature diagrams in phonology. In: Blackwell, A.F., Marriott, K., Shimojima, A. (eds.) Diagrams 2004. LNCS (LNAI), vol. 2980, pp. 398–401. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-25931-2_50
15. Sato, Y., Mineshima, K.: Human reasoning with proportional quantifiers and its support by diagrams. In: Jamnik, M., Uesaka, Y., Elzer Schwartz, S. (eds.) Diagrams 2016. LNCS (LNAI), vol. 9781, pp. 123–138. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42333-3_10

16. Sato, Y., Mineshima, K., Takemura, R.: The efficacy of Euler and Venn diagrams in deductive reasoning: empirical findings. In: Goel, A.K., Jamnik, M., Narayanan, N.H. (eds.) Diagrams 2010. LNCS (LNAI), vol. 6170, pp. 6–22. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14600-8_6
17. Scholz, B.C., Pelletier, F.J., Pullum, G.K.: Philosophy of linguistics. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Stanford University (2016)
18. Shimojima, A.: The graphic linguistic distinction. Artif. Intell. Rev. **13**(4), 313–335 (1999)
19. Stenning, K., Oberlander, J.: A cognitive theory of graphical and linguistic reasoning: logic and implementation. Cogn. Sci. **19**, 97–140 (1995)
20. Tversky, B., Corter, J.E., Yu, L., Mason, D.L., Nickerson, J.V.: Representing category and continuum: visualizing thought. In: Cox, P., Plimmer, B., Rodgers, P. (eds.) Diagrams 2012. LNCS (LNAI), vol. 7352, pp. 23–34. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31223-6_8
21. Wurman, R.S.: Information Architects. Graphis Press Corporation, New York (1996)

# Reasoning with Diagrams

# Accessible Reasoning with Diagrams: From Cognition to Automation

Zohreh Shams[1](✉), Yuri Sato[2], Mateja Jamnik[1], and Gem Stapleton[2]

[1] Department of Computer Science and Technology, University of Cambridge,
Cambridge, UK
{zohreh.shams,mateja.jamnik}@cl.cam.ac.uk
[2] Centre for Secure, Intelligent and Usable Systems, University of Brighton,
Brighton, UK
{y.sato,g.e.stapleton}@brighton.ac.uk

**Abstract.** High-tech systems are ubiquitous and often safety and security critical: reasoning about their correctness is paramount. Thus, precise modelling and formal reasoning are necessary in order to convey knowledge unambiguously and accurately. Whilst mathematical modelling adds great rigour, it is opaque to many stakeholders which leads to errors in data handling, delays in product release, for example. This is a major motivation for the development of diagrammatic approaches to formalisation and reasoning about models of knowledge. In this paper, we present an interactive theorem prover, called iCon, for a highly expressive diagrammatic logic that is capable of modelling OWL 2 ontologies and, thus, has practical relevance. Significantly, this work is the first to *design* diagrammatic inference rules using insights into what humans find accessible. Specifically, we conducted an experiment about relative cognitive benefits of *primitive* (small step) and *derived* (big step) inferences, and use the results to guide the implementation of inference rules in iCon.

## 1 Introduction

The long-held assumption that using diagrams makes modelling and reasoning accessible, goes back to ancient times (e.g., Euclid's Elements). Despite this, the development of automated diagrammatic reasoning tools (e.g., Hyperproof [2], Diamond [8] and Speedith [22]) has been rare in comparison to sentential theorem provers. One of the main areas yet to be explored in diagrammatic reasoning is the level of abstraction employed when constructing proofs, which relies on inference rule style. The 'right' level of abstraction can facilitate interaction between users and the system as well as increase the readability of the generated proofs [3,11]. But what is the 'right' level of abstraction for a human user? In this paper we introduce a diagrammatic reasoning system, iCon, for concept diagrams [19], which is designed so that the level of abstraction for diagrammatic inference rules is based on empirical results of what humans find accessible.

In sentential theorem proving, *tactics*, tacticals, proof strategies, proof methods and 'derived rules' are all attempts to achieve higher level of abstraction in

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 247–263, 2018.
https://doi.org/10.1007/978-3-319-91376-6_25

logical reasoning [5]: they enable applying a sequence of inference rules all in one go. Tools such as Isabelle [13] exploit tactical reasoning to provide a high level of abstraction and some level of automation. But the use of tactics in diagrammatic logics is largely unexplored. One attempt at controlling the level of abstraction in diagrammatic theorem proving is seen in Speedith [22] (a theorem prover for spider diagrams [4]) which uses tactics [11]. The choice of tactics in [11] is guided by metrics, which are informed by empirical studies [10] related to readability such as proof length and diagram clutter.

We focus on concept diagrams, which are based on spider diagrams but are more expressive. They were developed as a formal visualisation method for defining and reasoning about ontologies. Empirical evidence suggests [6] that concept diagrams are more accessible than the standard ontology language[1] OWL 2 and description logic [1]. In addition, cognitive theories support their effectiveness over common node-link ontology representation approaches [18].

Our goal is to develop a diagrammatic theorem prover for concept diagrams whose inference rules are designed to be accessible to users. We conducted an experiment to assess what level of abstraction users find accessible. We studied 10 inference tasks, based on practically relevant ontology inference problems formulated in [12]. The tasks were presented in two variations: using *primitive* inference steps and using *derived* inference steps, where the latter are coarser and more abstract than the former. User performance was measured in terms of their accuracy in identifying the validity of inference tasks. The level of abstraction (i.e., primitive vs. derived) that resulted in significantly higher accuracy rate was interpreted as the 'right' level of abstraction. We found that an appropriate level of abstraction was rule dependent. These results serve as the basis of design and implementation of inference rules in our interactive theorem prover iCon.

One of the unique selling points of iCon, and the main contribution of this paper, is the fact that the design of inference rules is guided by empirical studies of what inference rules people performed more accurately with, rather than being motivated by meta-level considerations such as completeness, which is often the case in formal systems. Although tactics incorporated in Speedith [22] are based on metrics that are found empirically, in this paper we have taken a step further by empirically testing the inferences themselves.

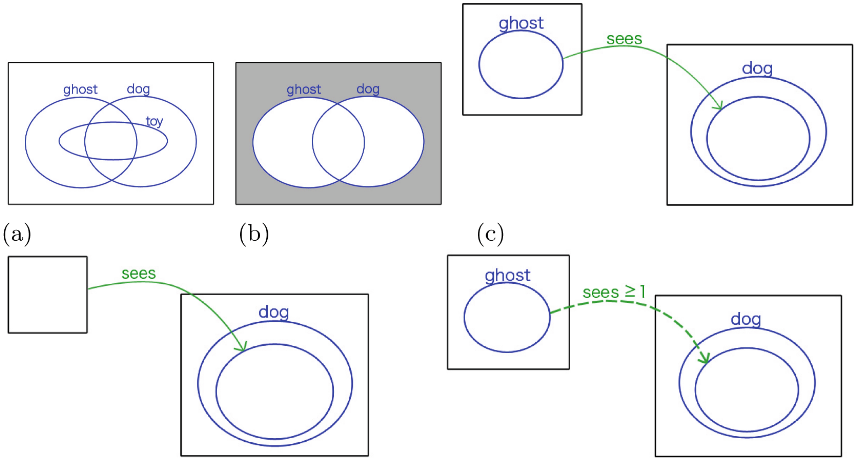
We give an overview of concept diagrams in Sect. 2, and introduce iCon in Sect. 3. In Sect. 4 we report on the empirical study that compares primitive vs. derived inference tasks. How the results of the empirical study inform the design of diagrammatic inference rules in iCon is discussed in Sect. 5. We compare our contributions to related work in Sect. 6 and finally, conclude in Sect. 7.

## 2 Concept Diagrams: Background and Overview

Concept diagrams were introduced for the purpose of visualising and specifying ontologies, and they are expressive enough to handle binary predicates [19]. They

---

[1] https://www.w3.org/TR/owl2-direct-semantics/.



(a) (b) (c)

(d) (e)

**Fig. 1.** Concept diagrams: (a) toys are either ghosts, dogs, or both; (b) there is nothing which is not a ghost or a dog; (c) all ghosts see *only* dogs; (d) all individuals in the world see *only* dogs; (e) all ghosts see *at least* one dog.

consist of syntactic objects such as rectangles, closed curves, and shading (as seen in Euler and Venn diagrams) as well as other additional objects such as dots, solid arrows and dashed arrows.

Rectangles are used to represent all individuals in the world. By combining curves inside a rectangle, we can represent several cases. For example, in Fig. 1(a), the circle toy is inside two circles, ghost and dog. Thus, toys are either ghosts, dogs, or both. Note that concept diagrams do not adopt the existential import assumption: the presence of a minimal region says nothing about whether there are some individuals in it (for details, see [15]). Shading is used to represent that there is nothing (i.e., to assert set emptiness). In Fig. 1(b), the region outside of ghost and dog is shaded. This means that there is nothing which is not a ghost or a dog. That is, everything is a ghost or a dog. The syntax described so far should be familiar in that concept diagrams with only rectangles, closed curves and shading are Euler diagrams.

Concept diagrams add syntactic objects to Euler diagrams, including arrows which are used to express verb relations. There are two kinds of arrows: solid and dashed ones. Solid arrows mean that the source is related to *only* the target. For example, in Fig. 1(c), the solid arrow labelled sees connects from the circle ghost to the unlabelled circle inside dog. This means that all ghosts see *only* dogs. Figure 1(d) is another example where the solid arrow sees connects from the rectangle to the unlabelled circle inside the circle dog. This means that all individuals in the world see *only* dogs. On the other hand, dashed arrows mean that the source is related to *at least* the target. That is, the source may be also related to other targets. For example, in Fig. 1(e), the dashed arrow sees connects the circle ghost to the unlabelled circle inside the circle dog. Together with the arrow annotation $\geq$1, this means that all ghosts see *at least* one dog.

Here is an informal overview of the syntax and semantics of concept diagrams; for formalisation, see [20]. A *concept diagram* is a collection of *boundary rectangles* including the syntax contained by them, and all arrows *connecting* them. Each boundary rectangle properly contains some (possibly empty) set of closed curves, some of which (possibly none or all) are labelled. Labelled closed curves represent specific sets (e.g., the set of dogs in Fig. 1(d)) or an anonymous set (e.g., some unnamed subset of dogs in Fig. 1(d)). The closed curves within a rectangle partition the plane into *zones*: a zone is a region inside some (possibly no) curves and outside the rest of the curves. For example, in Fig. 1(d), the concept diagram comprises two boundary rectangles, one of which contains no curves, the other contains two curves; these two curves give rise to three zones. In general, a set of zones inside a boundary rectangle is called a *region*.

Zones can be shaded and they may also contain dots. In addition, dots can be joined together by straight lines to form *spiders*. Each spider represents an individual. If the spider is labelled, it represents a specific individual. An unlabelled spider, just like an unlabelled curve, represents an anonymous individual. Distinct spiders represent distinct individuals, unless joined by = to assert their equality, or by $\overset{?}{=}$ to indicate that it is unknown whether they represent the same individual. Also, the individual represented by a spider is an element of the sets represented by the region in which the spider is placed.

The last major component of concept diagrams is arrows, which are of two types, dashed and solid. Arrows are sourced and targeted on boundary rectangles, closed curves, or spiders. Each arrow has a label, $p$, which represents a binary relation. The source and target of any given arrow need not be inside the same boundary rectangle. In addition, the label can be annotated with $^-$ to indicate the inverse of the relation, or with cardinality constraints: $\leq n$, $\geq n$ or $= n$, where $n$ is a natural number. Semantically, a solid arrow with source $s$, label $p$ (resp. $p^-$) and target $t$ expresses (blurring the distinction between syntax and semantics) that if the domain of $p$ (resp. $p^-$) is restricted to the source $s$ then the image is $t$. If, however, the arrow is instead dashed, then it expresses that if the domain of $p$ (resp. $p^-$) is restricted to the source $s$ then the image is a *superset* of $t$. In Fig. 1(c), the solid arrow expresses that, under the relation sees with domain restricted to ghost, the image is an anonymous subset of dog. Intuitively, this means that ghosts see only dogs. In Fig. 1(e), the dashed arrow expresses that, under the relation sees with domain restricted to ghosts, the image includes some anonymous subset of dog. Intuitively, this does not tell us anything. It is only through the use of the additional annotation, $\geq$1, that this arrow provides information: each ghost sees at least one dog.

Each rectangle, and its contents, in a concept diagram is called a *class and object property diagram*. This is because its curves represent classes (which are sets of individuals) and its arrows give information about properties (which are binary relations). Thus, a concept diagram is a set of class and object property diagrams, along with any connecting arrows.

# 3 iCon: A Concept Diagrams Interactive Theorem Prover

We built an interactive theorem prover[2] for concept diagrams, iCon, that can be used to reason, for example, about ontologies. iCon consists of the *reasoning engine* and the *graphical user interface (GUI).* We based iCon's design on Speedith [22], a theorem prover for spider diagrams [4], since concept diagrams are based on spider diagrams. In Speedith, the design of inference rules was based on obtaining soundness and completeness, whereas the design of inference rules in iCon is guided by an experiment (Sect. 4) into what abstraction level of deduction steps do people perform more accurately with. This approach to designing iCon improves the readability of the resulting proofs.

## 3.1 Reasoning Engine

The iCon reasoning engine (i) contains a collection of *inference rules*; (ii) handles the application of inference rules to diagrams expressed in an abstract syntax; and (iii) manages *proofs*. iCon only applies valid inference rules, and, since these are sound, proofs generated in iCon are guaranteed to be correct.

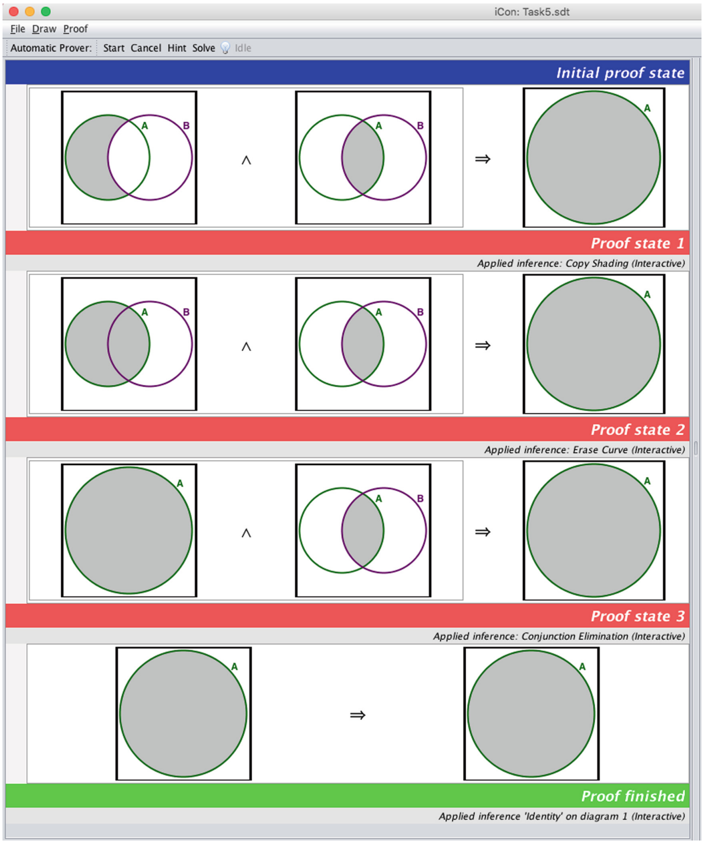
**Proofs.** A proof in iCon starts with the *initial proof state*, denoted as $\varDelta_0$ which is of form $(d_1 \wedge \cdots \wedge d_m) \Rightarrow d_n$, where $d_i$ are concept diagrams. This means that if $d_1, \cdots, d_m$ (referred to as *goals*, and denoted as set $G$) hold, then $d_n$ holds. Proofs are linear and constructed by applying sequences of *inference rules* on goals. Starting from $\varDelta_0$, the proof continues by applying inference rules to a goal $d \in G$. The result of applying an inference rule (with the exception of the inference rule *Identity* that will be explained in the next section) is a diagram $d'$, such that $d$ semantically entails $d'$ $(d \vDash d')$. In a proof, $P = \varDelta_0, \cdots, \varDelta_k$ s.t. $0 \leq i < k$, there has to be a goal $d$ in the set of goals in proof state $\varDelta_i$ and $d'$ in the set of goals in proof state $\varDelta_{i+1}$ such that $d'$ is the result of applying one of the inference rules to $d$. Proof $P$ is finished if the final proof state $\varDelta_k$ is of the form $d_n \Rightarrow d_n$, which means $d_n$ implies $d_n$, and is trivially true.

Figure 2 shows a proof where proof states are separated by red bars. The inference rules are applied on the proof states above the lines and result in the proof states below the lines (stated as "Applied inference"). In the final proof state, as expected, the diagram on the left hand side is identical to the one on the right hand side. Finally, the proof is finished by applying *Identity.* The rest of inference rules used in this figure will be formally defined in Sect. 5 (page 11).

**Inference Rule Design.** Inference rules are divided into logical and diagrammatic rules, where the former correspond to entailments and equivalences of propositional logic, while the latter rewrite the diagrams representing the goals. The fragment of concept diagrams used in this paper only uses $\wedge$ operator; therefore, the logical rules applicable are: *Conjunction Elimination* $((d_1 \wedge d_2) \Rightarrow d_1$

---

[2] Available at: https://github.com/ZohrehShams/iCon.



Initial proof state

Proof state 1

Applied inference: Copy Shading (Interactive)

Proof state 2

Applied inference: Erase Curve (Interactive)

Proof state 3

Applied inference: Conjunction Elimination (Interactive)

Proof finished

Applied inference 'Identity' on diagram 1 (Interactive)

**Fig. 2.** An example of a proof.

and $(d_1 \wedge d_2) \Rightarrow d_2$), *Conjunction Idempotency* ($(d \wedge d) \Leftrightarrow d$), and *Identity* (applied to the final proof state to express that $d \Rightarrow d$ is trivially true, and thus concludes the proof). The diagrammatic inference rules are motivated by the domain in which the reasoner is intended to be used: we chose ontology reasoning and debugging. We focus on a study [12] that provides statistical evidence for the practical significance, commonality and coverage of inference rules that it introduces for ontology entailment reasoning.[3] There are 51 inferences (referred to as ontology deduction patterns) identified in [12] and are ranked based on their accessibility measured through user studies. Each inference rule consists of up to four premises and a single conclusion and can often be broken down to more fine grained inference rules by introducing intermediate steps.

When designing the chosen diagrammatic inference rules, there are important choices with regards to the level of abstraction for these rules. To inform these choices, we conducted an empirical study (see Sect. 4) that compares two vari-

---

[3] Any ontology reasoning task can be reduced to entailment reasoning.

ations of diagrammatic inference rules, with different levels of abstraction and granularity in a number of deduction patterns. One variation takes the premises and conclusion of the deduction patterns in [12], and introduces an intermediate step; we call this variation *primitive*. The other variation is *identical* to the inference rules in [12], consisting of premises and conclusion only without an intermediate step; we call this variation *derived* – it is more abstract than the primitive version. The findings of the experiment comparing primitive and derived diagrammatic inference steps guides the process of implementing them, and it will be further discussed in Sect. 5.

### 3.2 Graphical User Interface

iCon provides a graphical user interface that allows visualising diagrams and applying inference rules on them interactively. Visualisation of diagrams is based on iCircles [21] – a Java library for drawing Euler diagrams using circles. iCircles was extended in Speedith [22] to represent spiders (existential elements). Here we extended it further to visualise unlabelled and labelled spiders and curves, and also to visualise solid and dashed labelled arrows with possible cardinalities. Users can select via a graphical point-and-click mechanism any part of the diagram to apply a diagrammatic inference rule on. This changes the diagram's abstract representation, and the new abstract representation is visualised using the visualiser. For screenshot of iCon, see Fig. 2.

## 4 Empirical Experiment for the Design of Inference Rules

To determine the right level of abstraction for diagrammatic rules, we compared participants' performance in inference tasks with concept diagrams in examples proved using primitive rules with those proved by derived inference rules.

### 4.1 Method

Fifty-one undergraduate students from seven classes on elementary computer science at the University of Brighton were recruited. The mean age was 24.12 ($SD = 5.89$) with a range of 19–48 years. All participants gave informed consent and were paid for their participation. The experiment method was approved by the CEM School Research Ethics Panel of University of Brighton. In order to provide an inference system that is accessible to a broad range of people, not just ontology experts, none of the participants had any prior knowledge of ontology engineering. One participant withdrew, so their data was excluded. Participants were randomly divided into two groups: the primitive group (N = 25) and the derived group (N = 25).

The participants in the primitive group were given tasks with intermediate step diagrams as well as premise and conclusion diagrams (i.e., primitive rules are applied in the proof; see Fig. 3 on page 9). The participants in the derived group were given tasks with only premise and conclusion diagrams without intermediate steps (i.e., derived rules are applied). Participants were asked to determine if

**Table 1.** List of inference tasks (#01–10 are valid; #11–20 are invalid. #01–05/#11–15 are Euler-Venn diagram level; #06–10/#16–20 are concept diagram level) and their accuracy rates ($*$ refers to a significant difference between the two groups at $p < 0.05$; $+$ refers to $p < 0.10$).

| Number | Premises ⇒ Conclusion | Primitive% | | Derived% |
|---|---|---|---|---|
| 01 | $(A \sqsubseteq B) \wedge (\mathit{Dis}(A, B)) \Rightarrow (A \sqsubseteq \bot)$ | 95.0 | | 80.1 |
| 02 | $(\mathit{Dis}(A, B)) \wedge (C \sqsubseteq A) \wedge (D \sqsubseteq B) \Rightarrow (\mathit{Dis}(C, D))$ | 80.0 | | 85.7 |
| 03 | $(A \sqsubseteq (B \sqsubseteq C)) \wedge (B \sqsubseteq C) \Rightarrow (A \sqsubseteq C)$ | 60.0 | $+$ | 85.7 |
| 04 | $(\top \sqsubseteq B) \wedge (\mathit{Dis}(A, B)) \Rightarrow (A \sqsubseteq \bot)$ | 60.0 | | 66.7 |
| 05 | $(A \sqsubseteq B) \wedge (A \sqsubseteq \neg B) \Rightarrow (A \sqsubseteq \bot)$ | 70.0 | | 76.2 |
| 06 | $(A \sqsubseteq \exists R.B) \wedge (\mathit{Rang}(R, C)) \Rightarrow (C \sqsubseteq \exists R.(B \sqcap C))$ | 75.0 | | 80.1 |
| 07 | $(A \sqsubseteq \exists R.(B \sqcap C)) \wedge (\mathit{Dis}(B, C)) \Rightarrow (A \sqsubseteq \bot)$ | 70.0 | | 85.7 |
| 08 | $(A \sqsubseteq \geq 3R.B) \wedge (A \sqsubseteq \leq 1R.B) \Rightarrow (A \sqsubseteq \bot)$ | 50.0 | | 52.4 |
| 09 | $(A \sqsubseteq \exists R.B) \wedge (B \sqsubseteq \bot) \Rightarrow (A \sqsubseteq \bot)$ | 65.0 | | 57.1 |
| 10 | $(A \sqsubseteq \geq 4R.B) \wedge (\mathit{Fun}(R)) \Rightarrow (A \sqsubseteq \bot)$ | 20.0 | $+$ | 47.6 |
| 11 | $(B \sqsubseteq A) \wedge (\mathit{Dis}(A, B)) \Rightarrow (A \sqsubseteq \bot)$ | 75.0 | | 71.4 |
| 12 | $(\mathit{Dis}(A, B)) \wedge (C \sqsubseteq A) \wedge (B \sqsubseteq D) \Rightarrow (\mathit{Dis}(C, D))$ | 70.0 | | 52.4 |
| 13 | $(A \sqsubseteq (B \sqsubseteq C)) \wedge (B \sqsubseteq C) \Rightarrow (A \sqsubseteq B)$ | 55.0 | | 47.6 |
| 14 | $(\top \sqsubseteq B) \wedge (A \sqsubseteq B) \Rightarrow (A \sqsubseteq \bot)$ | 60.0 | | 66.7 |
| 15 | $(A \sqsubseteq B) \wedge (\neg B \sqsubseteq A) \Rightarrow (A \sqsubseteq \bot)$ | 50.0 | $*$ | 85.7 |
| 16 | $(A \sqsubseteq \exists R.B) \wedge (\mathit{Rang}(R, C)) \Rightarrow (\mathit{Rang}(R, C \sqcap \neg B))$ | 40.0 | | 38.1 |
| 17 | $(A \sqsubseteq \exists R.(B \sqcup C)) \wedge (\mathit{Dis}(B, C)) \Rightarrow (A \sqsubseteq \bot)$ | 85.0 | | 85.7 |
| 18 | $(A \sqsubseteq \geq 1R.B) \wedge (A \sqsubseteq \leq 3R.B) \Rightarrow (A \sqsubseteq \bot)$ | 70.0 | | 66.7 |
| 19 | $(A \sqsubseteq \exists R.B) \wedge (A \sqsubseteq \bot) \Rightarrow (B \sqsubseteq \bot)$ | 60.0 | | 66.7 |
| 20 | $(A \sqsubseteq \geq 1R.B) \wedge (\mathit{Fun}(R)) \Rightarrow (A \sqsubseteq \bot)$ | 90.0 | | 81.0 |

the diagram transformations were valid or not. We presented 20 items in total: 10 consisted only of valid transformations of diagrams (#01–10) and 10 items included invalid transformations of diagrams (#11–20). The valid 10 items were selected from the *medium* difficulty amongst the 51 patterns given in [12]. This is because in any study, tasks that are too easy (leading to ceiling effect) or too hard (floor effect) reveal no insights. Minimal changes were made to the valid items (e.g., a set name, relation, cardinality) to create invalid ones. The tasks were further divided into the so-called Euler-Venn diagram level (#01–05; #11–15; they include only labelled curves and shading), and the concept diagram level (#06–10; #16–20; they include arrows and other syntax); the tasks are summarised using stylised description logic syntax in Table 1.[4] Tasks #01 and #06 are semantically equivalent and can be expressed by the same diagram-

---

[4] For unfamiliar readers, informally the DL syntax has the following interpretation: $A \sqsubseteq B$: $A$ is a subset of $B$; $\bot$: the empty set; $\exists R.A$: the set of things related to something in set $A$ under binary relation $R$; $\mathit{Rang}(R.A)$: the range of $R$ is $A$; $\mathit{Fun}(R)$: $R$ is functional; $\geq nR.A$: the set of things related to at least $n$ things in $A$ under $R$; $\leq nR.A$: the set of things related to at most $n$ things in $A$ under $R$.

matic representation. Thus, #01 was expressed by Venn diagrams and #06 was expressed by Euler diagrams. The tasks were presented in one of three random orders as a paper-and-pencil test. There was no time limit for completing the tasks, although the approximate time (30 min) for taking the experiment was instructed.

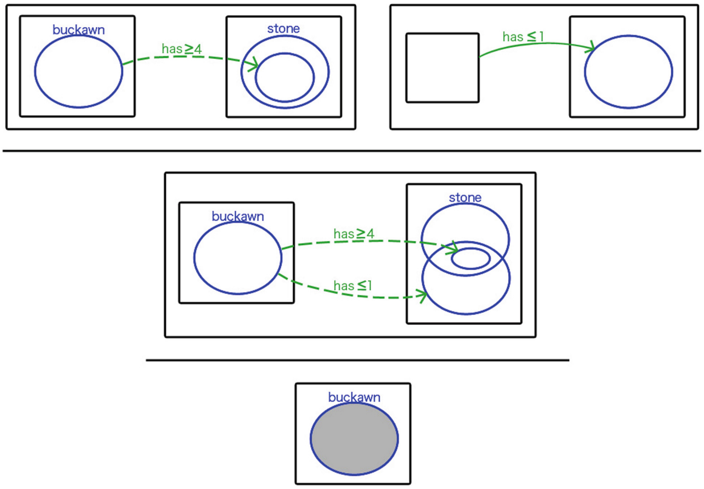


**Fig. 3.** A task (#10) in the primitive group. In the derived group, the intermediate diagram was removed.

All participants were gathered in a room. First, the participants were provided with three pages of instructions on the basic meaning of concept diagrams, but not on particular rules to solve inference tasks. Second, a pretest was conducted to check whether they understood the instructions correctly; they were asked to choose, from a list of three possibilities, the sentence corresponding to the meaning of a given diagram (for the importance of pretest settings, see [15]). Third, the participants were provided with one page of instruction on the meaning of a valid transformation (entailment), with two examples of diagrams: one was valid and one was not valid.[5] After the instruction phase, the participants were asked to solve the main reasoning tasks of the experiment.

## 4.2 Results

The data for participants who made mistakes in more than two items (out of five) in the pretest was removed. In our analysis, 5 out of 25 in the primitive group, and 4 out of 25 participants in the derived group were removed.

---

[5] See https://sites.google.com/site/myardproject/exp/MateInst2.zip?aredirects=0&d=1 for full instructions.

The accuracy data for each task was analysed using a $\chi^2$ test. In task #15, accuracy rates in the derived group were significantly higher than those in the primitive group (47.1% vs. 88.2%, $p = 0.014$). In task #03, accuracy rates in the derived group were significantly higher than those in the primitive group, at a reduced threshold of $p < 0.10$ (64.7% vs. 88.2%, $p = 0.063$). In task #10, accuracy rates in the derived group were significantly higher than those in the primitive group, at a reduced threshold of $p < 0.10$ (23.5% vs. 47.1%, $p = 0.062$). In other tasks, there were no significant differences between both groups.

In the comparison between #01 (expressed with Venn diagrams) and #05 (expressed with Euler diagrams), a significant difference was found in the primitive group (95.0% vs. 70.0%, $p = 0.037$), but not in the derived group (80.1% vs. 76.2%, $p = 0.432$). Overall, the comparison of accuracy data between the primitive group and the derived group revealed no significant difference.[6]

### 4.3 Discussion

In task #10, shown in Fig. 3 (described as $(A \sqsubseteq \geq 4R.B) \wedge (Fun(R)) \Rightarrow (A \sqsubseteq \bot)$), 80% of participants *incorrectly* judged the validity of diagram transformation using primitive rules. In comparison to the (random) chance level of 50%, there was a significant difference in the primitive group ($p = 0.047$), but not in the derived group ($p = 0.877$). In the diagram transformation using primitive rules, the solid arrow from the rectangle is explicitly rewritten into the dashed arrow from the curve inside the rectangle. On the other hand, there is no explicit rewriting of solid and dashed arrows in the diagram transformation using derived rules. Since this difference is found in task #10, it is a candidate to explain the result of task #10 where better accuracy was achieved with the derived rule. Therefore, a resulting heuristic suggests that we should not design diagram transformations where solid arrows are replaced by dashed arrows.

On the other hand, what causes the significant differences in tasks #03 and #15 between the primitive and the derived groups? In the diagram transformation in task #03, as shown in Fig. 4 (left), it is important that the curves crossing between lizardfolk and kobold in the first premise diagram mean that the semantic relationship between them is indeterminate (see the existence-free assumption for minimal regions, mentioned in Sect. 2). Therefore, the unification between the premise diagrams results in the diagram where (i) lizardfolk is inside kobold, (ii) merfolk is inside kobold, (iii) the relation between lizardfolk and merfolk is unknown. Here, understanding crossing curves plays an essential role in both deducing (i–iii) using a primitive rule *and* deducing (ii) using a derived

---

[6] 65.0% vs. 69.1% (Mann-Whitney $U = 179$, $p = 0.416$) for overall (#01–20), 64.5% vs. 71.9% ($U = 179$, $p = 0.159$) for valid transformations (#01–10), 65.5% vs. 66.2% ($U = 208$, $p = 0.958$) for invalid ones (#11–20), 73.0% vs. 79.1% ($U = 166$, $p = 0.232$) for valid Euler ones (#01–05), 56.0% vs. 64.8% ($U = 166.5$, $p = 0.242$) for valid concept ones (#06–10), 52.0% vs. 64.8% ($U = 199$, $p = 0.769$) for invalid Euler ones (#11–15), and 69.0% vs. 67.6% ($U = 194.5$, $p = 0.673$) for invalid concept ones (#15–20).

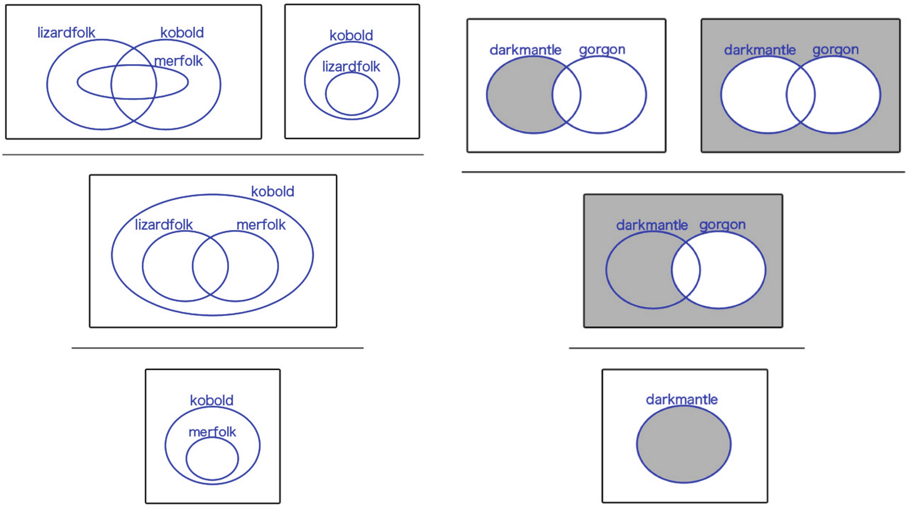


**Fig. 4.** Task #03 (left) and task #15 (right) in the primitive group. In the derived group, the intermediate diagrams were removed.

rule. Thus, the meaning of crossing curves cannot explain the performance difference between both groups. In the case of task #15, as shown in Fig. 4 (right), shading plays an important role in solving the task. However, the same can be said for tasks #01, 04, 11, and 14, where significant differences in accuracy performances were not found. Thus, it is not clear why the difference between both groups was found only in task #15.

As stated before, tasks #01 and #05, which are semantically equivalent, were expressed by Venn and Euler diagrams, respectively, as shown in Fig. 5. The result that the accuracy rate for #01 was higher than for #05 in the primitive group suggests that Venn diagrams are more suitable than Euler for reasoning about the emptiness of a set.[7] In #05 (using Euler diagrams), the derivation of the shaded curve labelled darkmantle, meaning darkmantle $\sqsubseteq \bot$, requires not only spatial operation on diagrammatic objects, but also meta-level information concerning semantic values (cf. the discussion in [16]). Whether for primitive rules or derived rules, reasoning with Euler diagrams in this case can require more cognitive effort than with Venn diagrams. Note that the effectiveness of Venn over Euler diagrams is distinct from previous empirical findings [15].

## 5 Inference Rules: Design and Implementation Guidelines

We chose ontology reasoning and debugging for the first application domain of iCon. In order to develop a theorem prover with practical relevance in this

---

[7] Note that in ontology engineering, sets that are necessarily empty are called *incoherent*.

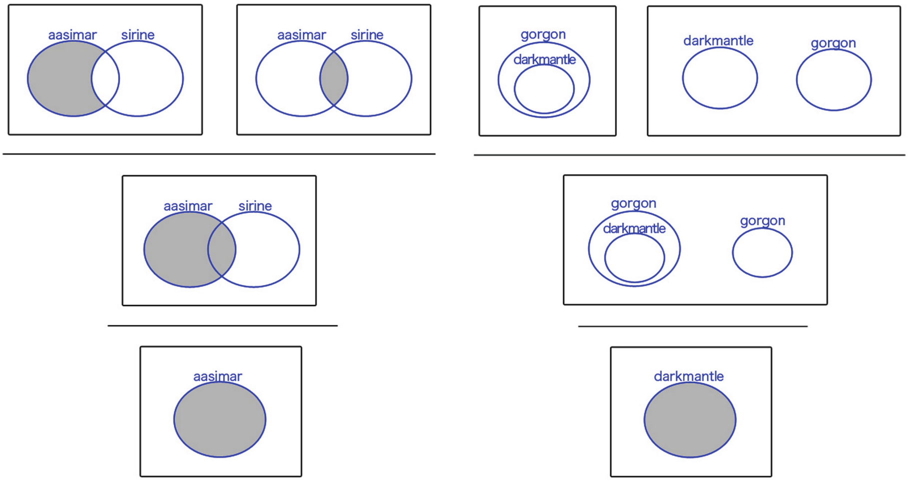


**Fig. 5.** Task #01 (left) and task #05 (right) in the primitive group. In the derived group, the intermediate diagrams were removed.

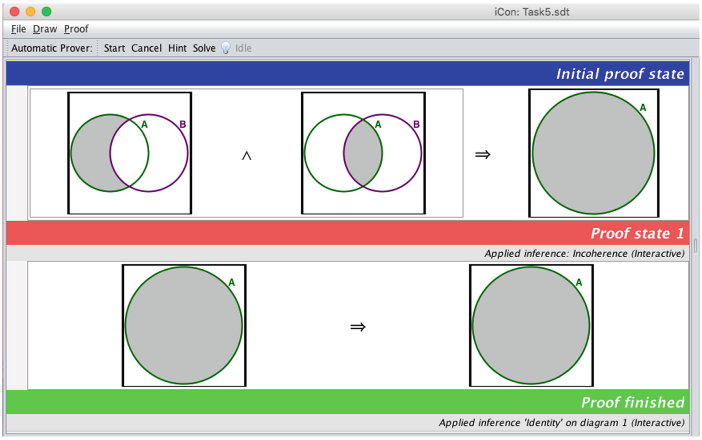
domain, we focused on the 51 inferences found in [12]. However, there are a lot of choices when designing the diagrammatic version of these inferences in iCon. For example, the inference patterns are large and can often be broken down into smaller steps, but what is the right level of granularity for the diagrammatic inference rules? To ensure accessibility, this granularity level was informed by our experiment where we translate the results into design and implementation guidelines. Note that in the user study we tested the accessibility of 10 inferences out of 51 from [12], but the design guidelines are general and can be used in the implementation of any concept diagram inference rules.

Overall, no significant difference was observed between primitive and derived rules, which suggests that the level of abstraction in the implementation is a choice that the users ought to have and use as they see appropriate. Thus, the first guideline we extract from the experiment is to implement both primitive and derived versions of the inference rules.

The second guideline is related to the heuristic (Sect. 4.3, first paragraph) that suggests that inference rules should not transform solid arrows into dashed arrows. In the user study, tasks #6–10 involve arrows, with only task #10 transforming arrows in its primitive version (Fig. 3). Thus, we adopted the primitive version of task #10, such that in the intermediate step diagram, the arrow with cardinality $\leq 1$ stays solid (right hand premise in Fig. 3). For the remaining 41 inferences, we employ this heuristic and ensure that their diagrammatic versions retain the arrow type.

The third guideline is based on the heuristic (Sect. 4.3, last paragraph) that Venn seems to be a more effective representation than Euler when proving incoherence of a set. More than 20% of 51 inferences from [12] prove incoherence, and thus should employ this heuristic. Concretely, in our study, tasks #1,#4, #5, and #7–10 deduce that a set is incoherent. Out of these, tasks #1,#4, #5, and #7 can be alternatively represented in Euler or Venn form. In the experiment, apart from task #5, which used Euler form (see Fig. 5 (right)), the rest of them were presented in Venn form. We revised task #5 accordingly, such that it is presented in Venn form.

We now exemplify how these guidelines are put into practice in iCon. We focus on the design and implementation of inference rules for task #5 in Fig. 5 (right). Following the first design guideline, inference rules for both, primitive and derived version of this task are implemented. In fact, the proof in Fig. 2 (page 6) shows the implementation of the primitive version, while Fig. 6 shows the implemented derived version, where $A$ is darkmantle and $B$ is gorgon. The second design guideline does not apply here, as there are no arrows. As seen in Figs. 2 and 6, following the third guideline presents the premises in Venn instead of the Euler form that was originally used in our study. We now formally define of the diagrammatic inference rules used in Figs. 2 and 6. These rules can be proved sound as in [17]. In addition to diagrammatic rules, the two logical rules used are *Conjunction Elimination* and *Identity* (see Sect. 3.1).



Applied inference: Incoherence (Interactive)

Applied inference 'Identity' on diagram 1 (Interactive)

**Fig. 6.** Proof of task # 5 using derived rule.

The first diagrammatic rule we define copies shading from one region to another. It relies on syntactically identifying when two regions represent the same set. This has been extensively studied for Euler diagrams [7] and spider diagrams [22]. Thus, the syntactic identification of regions that represent the same sets can be identified using the underlying Euler diagram (i.e., the boundary rectangle containing only the labelled curves). Given a set of labelled curves, $C$, a *fixed zonal region* is the set of zones that are inside all of the curves in $C$ and possibly other curves; note the name 'fixed' since such regions represent particular, that is, 'fixed' sets and are not anonymous. Given a region, $r$, in a concept diagram, $d$, we say that $r$ is *fixed* if it is formed from a union of fixed zonal regions. Fixed regions are said to be *corresponding* if, informally, they represent the same set (the details of how this can be identified syntactically can be found in [22]). We can now define the *Copy Shading* inference rule.

**Definition 1 (Copy Shading).** *Let* $d_1$ *and* $d_2$ *be two concept diagrams containing corresponding fixed regions* $r_1$ *and* $r_2$*, respectively, where:*

1. $r_1$ *comprises only shaded zones and* $r_2$ *has at least one non-shaded zone,*
2. *any spider with a dot in* $r_1$ *(resp.* $r_2$*) is completely contained by zones in* $r_1$ *(resp.* $r_2$*), and*
3. *the spiders in* $r_1$ *match the spiders in* $r_2$*.*

*Let* $d_2'$ *be a copy of* $d_2$ *except that* $r_2$ *is entirely shaded. From* $d_1 \wedge d_2$ *we can infer* $d_1 \wedge d_2'$ *and vice versa.*

The Next rule needed for task #5 deletes a curve from a concept diagram.

**Definition 2 (Erase Curve).** *Let* $c$ *be a curve in a concept diagram* $d$*. Then* $c$ *can be removed from* $d$*, resulting in a new diagram,* $d'$*, with modified shaded zones, spider habitats and arrows. In particular, if upon erasure of* $c$*, a shaded zone merges with a non-shaded zone then the shading is removed, otherwise the shading is preserved. Also, if a spider* $s$ *has a foot in two zones that collapse into one, then the spider will have a foot in the collapsed zone in* $d'$*. In addition, arrows that have* $c$ *as target or source are deleted when forming* $d'$*. From* $d$ *we can infer* $d'$*.*

Unlike the *Copy Shading* rule, which preserves semantics and is an equivalence, *Erase Curve* weakens information and can be applied only in one direction.

Our next rule, *Incoherence* is used in the derived version of task #5. It allows deducing that a curve, say $c$, is entirely shaded, by copying shading from a conjunct diagram in which the corresponding non-shaded region of $c$ are shaded.

**Definition 3 (Incoherence).** *Let* $d_1$ *and* $d_2$ *be two concept diagrams containing curves* $c_1$ *and* $c_2$*, respectively, such that:* $c_1$ *and* $c_2$ *have the same label as each other;* $c_1$ *and* $c_2$ *do not contain any spiders; and the non-shaded zones inside* $c_1$ *in* $d_1$ *form a fixed region that corresponds to some entirely shaded, fixed region contained by* $c_2$ *in* $d_2$*. Let* $d_3$ *be a concept diagram comprising a single boundary rectangle containing a curve,* $c_3$*, with the same label as* $c_1$ *and* $c_2$ *whose interior is entirely shaded. From* $d_1 \wedge d_2$ *we can infer* $d_3$*.*

Currently, iCon implements 7 out of 10 inferences from [12] that were used in the user study, plus additional rules that enable the user to vary the granularity of the proof. To provide the same coverage as in [12], we are currently building both primitive and derived variations for the remaining inferences in [12].

## 6 Related Work and Evaluation

Like iCon, DIAMOND [8] and Cinderella [9] are diagrammatic theorem provers, however they operate in different domains of inductive theorems of natural numbers and geometry, respectively. In contrast, iCon deals with theorems about sets. Unlike DIAMOND and iCon, proof steps in Cinderella may not be sound and have to be verified externally by an automatic symbolic theorem prover.

iCon's concept diagrams are a more expressive extension of Speedith's spider diagrams. Similarly to Speedith, the inference rules in iCon are purely logical or diagrammatic. However, in Speedith the choice of inference rules is motivated by the completeness property. In contrast, in iCon the focus is on the commonality of the inference rules in the ontology domain. In addition, the design and implementation of inference rules in iCon is informed by empirical studies of what people find intuitive, and in particular, what level of granularity of rules enables human users to reason most accurately. This is in line with one of the most challenging areas of theorem proving, which is reducing the gap between user's model of the proof and the actual proof constructed by mechanised theorem provers [3]. The proof steps in the user's model are often coarser and have intuitive semantics, whereas the prover's steps tend to be much more fine grained. Our user study presented inference rules at different levels of granularity. The derived ones, which are coarser than the primitive ones, can be seen as tactics, as is common in sentential theorem proving. However, the role of tactics in sentential theorem proving is typically to provide some level of automation, and rarely to reduce the gap between user's reasoning and that of a prover. Our work addresses both, automation as well as human approach to constructing proofs. For instance, in Fig. 6 the tactic *Incoherence* reduces the length of proof in comparison with Fig. 2, while it still remains accessible by allowing the user to choose a curve and establish its incoherence.

Speedith deploys diagrammatic tactics to facilitate user interactions and devise a higher abstraction level in proofs which renders them more self-explanatory [11]. The choice of tactics is guided by metrics related to the readability of proofs (e.g. length of proof, the amount of clutter) and are informed by empirical studies. Our work presents a step change in that we directly test the accessibility of inferences themselves.

## 7 Conclusion and Future Work

Usability and accessibility of reasoning systems is paramount to harness their utility in diverse domains. By developing an interactive diagrammatic theorem prover iCon, we demonstrated that it is possible to build a formal reasoning system that is based on empirical studies of what humans find accessible.

iCon implements the logic of concept diagrams and can be applied in various domains (e.g., for reasoning in ontology engineering as presented here). We focused on deduction patterns found in [12], which also discusses their significance in terms of commonality and coverage. In order to gain an insight into how to implement the concept diagrams version of these patterns, we conducted an empirical study that identified that the level of abstraction and granularity of inference rules did not affect what human reasoners find accessible in general, but was rule specific. We used this result and others explained in Sect. 4.3 to guide the design and implementation of iCon's inference rules.

Displaying the application of inferences via the GUI presents many challenges and avenues for future work. Laying out the drawn diagrams after each inference requires analysing the invariant parts of the diagrammatic statement, because these are the syntactic elements that must remain unchanged before and after the application of the inference rule. But there are choices and trade-offs between what elements could or should be preserved. We are planning to conduct a user study that will investigate where this trade-off lies with the human users. Furthermore, the layout algorithms of iCon should preserve certain wellformedness properties of the diagrams [14]. For example, ideally a curve should not be split into two disjoint curves with the same label. Improving layout algorithms for iCon's GUI remains future work.

**Acknowledgements.** This research was funded by a Leverhulme Trust Research Project Grant (RPG-2016-082) for the project entitled Accessible Reasoning with Diagrams. The authors would like to thank Prof. John Howse, Dr Andrew Blake and Dr Ryo Takemura for their cooperation in the experiments.

## References

1. Baader, F., Horrocks, I., Sattler, U.: Description logics. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, pp. 21–43. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-92673-3_1
2. Barwise, J., Etchemendy, J.: Hyperproof. CSLI Publications, California (1994)
3. Beckert, B., Grebing, S., Böhl, F.: A usability evaluation of interactive theorem provers using focus groups. In: Canal, C., Idani, A. (eds.) SEFM 2014. LNCS, vol. 8938, pp. 3–19. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-15201-1_1
4. Gil, J., Howse, J., Kent, S.: Formalizing spider diagrams. In: IEEE Symposium on Visual Languages, pp. 130–137. IEEE (1999). https://doi.org/10.1109/VL.1999.795884
5. Harrison, J., Urban, J., Wiedijk, F.: History of interactive theorem proving. In: Handbook of the History of Logic. Vol. 9: Computational Logic, pp. 135–214. Elsevier (2014)
6. Hou, T., Chapman, P., Blake, A.: Antipattern comprehension: an empirical evaluation. In: Formal Ontology in Information Systems. Frontiers in Artificial Intelligence, vol. 283, pp. 211–224. IOS Press (2016). https://doi.org/10.3233/978-1-61499-660-6-211
7. Howse, J., Stapleton, G., Flower, J., Taylor, J.: Corresponding regions in Euler diagrams. In: Hegarty, M., Meyer, B., Narayanan, N.H. (eds.) Diagrams 2002. LNCS (LNAI), vol. 2317, pp. 76–90. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-46037-3_7
8. Jamnik, M.: Mathematical Reasoning with Diagrams. CSLI Publications, California (2001)

9. Kortenkamp, U., Richter-Gebert, J.: Using automatic theorem proving to improve the usability of geometry software. In: Mathematical User-Interfaces Workshop (2004)
10. Linker, S., Burton, J., Blake, A.: Measuring user comprehension of inference rules in Euler diagrams. In: Jamnik, M., Uesaka, Y., Elzer Schwartz, S. (eds.) Diagrams 2016. LNCS (LNAI), vol. 9781, pp. 32–39. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42333-3_3
11. Linker, S., Burton, J., Jamnik, M.: Tactical diagrammatic reasoning. In: International Workshop on User Interfaces for Theorem Provers. Electronic Proceedings in Theoretical Computer Science, vol. 239, pp. 29–42 (2016). https://doi.org/10.4204/EPTCS.239.3
12. Nguyen, T.A.T., Power, R., Piwek, P., Williams, S.: Measuring the understandability of deduction rules for OWL. In: International Workshop on Debugging Ontologies and Ontology Mappings, pp. 1–12. Linköping Electronic Conference Proceedings (2012)
13. Paulson, L.C.: Isabelle - A Generic Theorem Prover (with a contribution by T. Nipkow), vol. 828. Springer, Heidelberg (1994). https://doi.org/10.1007/BFb0030541
14. Rodgers, P., Zhang, L., Purchase, H.: Wellformedness properties in Euler diagrams: which should be used? IEEE Trans. Vis. Comput. Graph. **18**(7), 1089–1100 (2012). https://doi.org/10.1109/TVCG.2011.143
15. Sato, Y., Mineshima, K.: How diagrams can support syllogistic reasoning: an experimental study. J. Log. Lang. Inf. **24**(4), 409–455 (2015). https://doi.org/10.1007/s10849-015-9225-4
16. Sato, Y., Ueda, K., Wajima, Y.: Strategy analysis of non-consequence inference with Euler diagrams. J. Log. Lang. Inf. **27**, 61–77 (2017). https://doi.org/10.1007/s10849-017-9259-x
17. Shams, Z., Jamnik, M., Stapleton, G., Sato, Y.: Reasoning with concept diagrams about antipatterns in ontologies. In: Geuvers, H., England, M., Hasan, O., Rabe, F., Teschke, O. (eds.) CICM 2017. LNCS (LNAI), vol. 10383, pp. 255–271. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62075-6_18
18. Shimojima, A.: Semantic Properties of Diagrams and Their Cognitive Potentials. CSLI Publications, California (2015)
19. Stapleton, G., Compton, M., Howse, J.: Visualizing OWL 2 using diagrams. In: IEEE Symposium on Visual Languages and Human-Centric Computing, pp. 245–253. IEEE (2017). https://doi.org/10.1109/VLHCC.2017.8103474
20. Stapleton, G., Howse, J., Chapman, P., Delaney, A., Burton, J., Oliver, I.: Formalizing concept diagrams. In: Visual Languages and Computing, pp. 182–187. Knowledge Systems Institute (2013)
21. Stapleton, G., Zhang, L., Howse, J., Rodgers, P.: Drawing Euler diagrams with circles: the theory of piercings. IEEE Trans. Vis. Comput. Graph. **17**(7), 1020–1032 (2011). https://doi.org/10.1109/TVCG.2010.119
22. Urbas, M., Jamnik, M., Stapleton, G.: Speedith: a reasoner for spider diagrams. J. Log. Lang. Inf. **24**(4), 487–540 (2015). https://doi.org/10.1007/s10849-015-9229-0

# Using Diagrams to Reason About Biological Mechanisms

William Bechtel[1(✉)], Adele Abrahamsen[2], and Benjamin Sheredos[3]

[1] Department of Philosophy, University of California,
San Diego, La Jolla, CA 92093-0119, USA
bechtel@ucsd.edu
[2] Center for Research in Language, University of California,
San Diego, La Jolla, CA 92093, USA
aabrahamsen@ucsd.edu
[3] Analytical Writing Program, University of California,
San Diego, La Jolla, CA 92093, USA
ben.sheredos@gmail.com

**Abstract.** In developing mechanistic explanations for biological phenomena, researchers have their choice of several different types of diagrams. First, a *mechanism diagram* spatially represents a proposed mechanism, typically using simple shapes for its parts and arrows for their operations. Beyond this representational role, such diagrams can provide a platform for further reasoning. Published diagrams in circadian biology show how question marks support reasoning about the proposed molecular mechanisms by flagging where there are knowledge gaps or uncertainties. Second, an *annotated mechanism diagram* can support computational modeling of the dynamics of a proposed mechanism. Each variable and parameter needed for the model is added to the diagram adjacent to the appropriate part or operation. Anchoring the model in this way helps with its construction, revision, and interpretation. Third, a *network diagram* fosters a different approach to mechanistic reasoning. Layout algorithms are applied to data generated by high-throughput experiments to reveal modules that correspond to mechanisms. We present examples in which network diagrams enable viewers to advance hypotheses about previously unknown mechanisms or unknown parts and operations of known mechanisms as well as to develop new understanding about how a given mechanism is situated in a larger environment.

**Keywords:** Mechanistic explanation · Mechanism diagrams
Network diagrams · Question marks · Computational model
Biological mechanism · Circadian rhythms

## 1 Introduction

Much research in the life sciences is devoted to developing mechanistic explanations of specific biological phenomena such as circadian rhythms and cell division. What distinguishes mechanistic explanations is that they associate a delineated phenomenon with an identified mechanism and propose a decomposition of that mechanism into

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 264–279, 2018.
https://doi.org/10.1007/978-3-319-91376-6_26

parts and operations that, when appropriately organized, are claimed to produce the phenomenon of interest [3, 14]. A mechanistic explanation can be represented spatially in what we call a *mechanism diagram,* usually using simple shapes (and/or words) for the parts of the proposed mechanism and arrows for their operations. Tversky [22, 23] refers to these shapes and arrows as *morphoglyphs* or *glyphs.* As she emphasizes, "Like words in language, morphoglyphs can be combined in various ways to create varying meanings." The diagrams constructed by combining these elements often appear as the final figure in research articles, where they show how the findings have yielded a new or improved understanding of the responsible mechanism. Mechanism diagrams play a more prominent role in review papers, textbooks, and lectures, where displaying contrasting mechanism diagrams is an effective way to compare alternative explanations or to show how understanding has changed over time. It should be noted that published diagrams are the final product of an often extensive process of development, a process we examined using one laboratory's unpublished drafts of the diagrams they eventually published [4, 20]. Here we focus on final, published diagrams and argue that these are offered not just as conclusions, but also as a basis for further reasoning.

The diagram research community is multidisciplinary. Our own way of investigating diagrams is grounded in the new mechanistic philosophy of science. [2, 3, 6] Those taking this approach have given considerable attention to the strategies by which scientists develop mechanistic explanations, and some have noted that scientists often include mechanism diagrams in their publications. With few exceptions, though, mechanists have not focused on how scientists rely on diagrams in their own reasoning about mechanisms.

Cognitive science is a field with the potential to pursue this question via experiments and other empirical methods. In fact, though, most cognitive scientists who study diagrammatic reasoning have tended to focus on relatively simple diagrams constructed or interpreted by participants in brief experiments, not the complex diagrams that have been painstakingly developed and modified by scientists or science educators. There are a few exceptions. Nersessian [16] drew novel insights about reasoning about physical forces from a close examination of Faraday's and Maxwell's diagrams, and since then has extended her work into systems biology. Hegarty [21], with an ongoing focus on reasoning with diagrams, recently examined the molecular diagrams of organic chemistry and how students learn to use them. Cheng [5] has explored new diagrammatic formats that support better learning about probability and electricity. There should be more such studies, but also initiation of research on how mechanism diagrams in particular support reasoning in biology and other sciences.

To whet the appetite of diagram researchers in philosophy, cognitive science, and other fields, we introduce three types of diagrams and their role in biological reasoning. In Sect. 2 we discuss mechanism diagrams that include multiple question marks to signify where further research is needed to fill a lacuna or to provide better supporting evidence. In Sect. 3, we show how a mechanism diagram is linked to a computational model of the mechanism's dynamics by annotating it with variables and parameters from the model. In Sect. 4 we turn to a different type of diagram that increasingly is being used to make inferences about mechanisms: network diagrams. We show how such diagrams are employed to formulate new hypotheses about parts and operations of mechanisms or how mechanisms are situated in larger environments.

## 2 Using Question Marks in Mechanism Diagrams

Question marks occur with surprising frequency in published mechanism diagrams. They are used to indicate knowledge gaps: uncertainty about the existence or identity of parts or operations needed to achieve a working mechanism. By visually pinpointing where further research is needed, the question marks in a mechanism diagram highlight that it cannot be regarded as a finished explanation but rather captures one moment in a process of inquiry. Here we discuss three cases illustrating how question marks can aid biologists' reasoning as they engage in that process.

First, a major paper by two recipients of the 2017 Nobel Prize in Physiology and Medicine includes a simple mechanism diagram with no fewer than five question marks (Fig. 1). The prize was awarded for research leading to a molecular mechanism hypothesized to be responsible for circadian rhythms. Two of the Laureates, Hall and Rosbash, were selected for their work identifying the role of the gene *period*. They were not the discovers of the gene or its crucial role in circadian rhythms—it was already recognized as crucial to circadian rhythms through the research of Konipka and Banzer [13], who induced mutations in fruit flies and screened for effects on circadian rhythms. Mutations in the gene they named *period* (abbreviated *per*) resulted in a circadian period that was shorter or longer than 24 h or was absent. What Hall and Rosbash's team added was cloning of the *per* gene to enable quantitative analysis of its expression. Crucially, they found that the cyclic rise in concentration of *per* mRNA preceded that of the PER protein by approximately four hours. Based on these findings, Hardin et al. [8] proposed the molecular mechanism shown in Fig. 1, in which PER proteins feed back into the nucleus to inhibit the transcription of their own gene (*per*). (For understanding Fig. 1 and other figures, it is helpful to know that typically gene names are italicized in lower case and protein names are either capitalized or written entirely in upper case.)

One important use of a mechanism diagram is to support reasoning about how the mechanism is capable of generating the phenomenon of interest. In particular, the diagram in Fig. 1 helps with mentally animating the functioning of the *per* feedback mechanism. Hegarty [10] presented undergraduates with a reasoning task based on a diagram of a different mechanism—a pulley system—to study this animation process. She found convincing evidence that her participants solved problems by mentally progressing stepwise through the diagram, attending to each operation in turn. Researchers benefit from diagrams of circadian mechanisms in the same way. Assume that PER is initially in low concentration. Then the operations in the upper left of Fig. 1 proceed as shown: first the transcription of *per* into its mRNA and then translation into the protein PER. But as PER reaches higher concentrations, negative feedback slows down those operations. Although not shown in the diagram (the reader was expected to supply such knowledge), PER molecules break down over time, gradually releasing *per* from inhibition so that transcription and translation ramp up again. The overall result is a cycle in which the abundance of PER waxes and wanes every 24 h.

While the basic idea of a feedback loop is clear enough and easy to represent in a diagram, Hardin et al. did not know where the feedback originated and whether it inhibited only translation into mRNA or transcription into the PER protein. When they

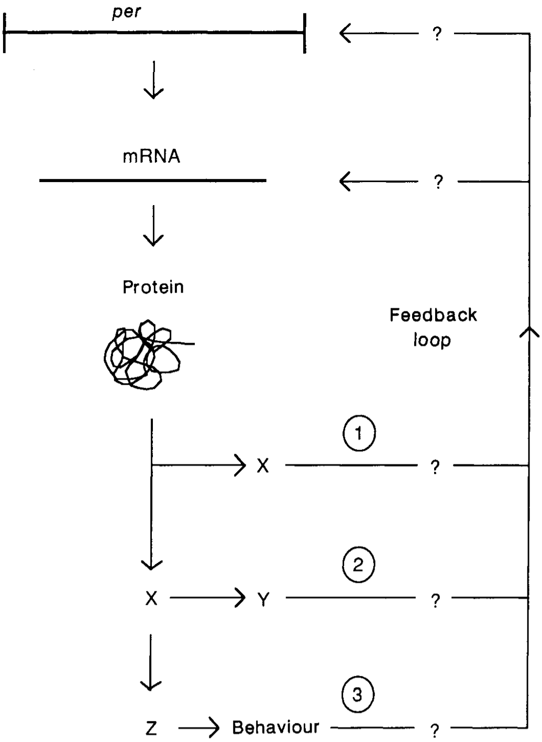


**Fig. 1.** Use of question marks to identify alternative hypotheses for the feedback from a protein X (no mediator or via protein Y or via protein Z and behavior) and for whether the feedback targets the protein's gene and/or its mRNA. Reprinted by permission from Nature/Springer/Palgrave, from [8].

presented their proposed mechanism diagrammatically, therefore, they made liberal use of question marks. For the uncertain source of the feedback they showed three options (numbered 1–3 in Fig. 1), each flagged by a question mark: (1) PER itself (X), (2) another protein (Y) that PER acted on, or (3) a behavior of the organism affected by PER (mediated by a protein Z). They also inserted question marks into the upper right arrows to indicate uncertainty as to which operations were affected by the feedback. Despite the uncertainties, the basic idea of a transcription-translation feedback loop (TTFL) was rapidly adopted and has guided circadian research in most species in the subsequent decades.

The nascent TTFL in Fig. 1, and the evidence that some version of this mechanism functioned as a circadian clock, was a major achievement. But far from signaling the end of inquiry, the question marks in it directed the community to further research that would require many years to complete.

A diagram proposing a mechanism, when included in a research paper, typically appears as the last figure (following a number of other figures displaying data). Hardin et al.'s diagram exemplifies this future-facing placement. But sometimes an unanswered question about a mechanism was what motivated a research project, in which case a mechanism diagram with questions marks may be the *first* figure in the published paper. Figure 2 provides an example. It is the first figure in Paddock et al. [17], a paper seeking to identify how the circadian clock in cyanobacteria regulates the expression of most other genes in that organism. Unlike the circadian clocks for all other orders of life, this bacterial clock depends not on proteins suppressing the expression of a gene but instead a cycle of reactions phosphorylating and dephosphorylating the protein KaiC. KaiC is shown in Fig. 2 as a lumpy double doughnut—a shape conveying that it contains two domains (labeled CI and CII on the left) and functions as a hexamer. It cycles through four different phosphorylation states: the black circles labeled P represent phosphates bound to site S (pST), site T (SpT), both (pSpT), or neither (ST). In each state, KaiC also has a different relation to the proteins KaiA and KaiB and, most important here, has a dashed arrow pointing to the word *Output* followed by a question mark. Paddock et al. use this device to ask: which one of the four phosphorylation states generates the output that helps regulate other mechanisms?

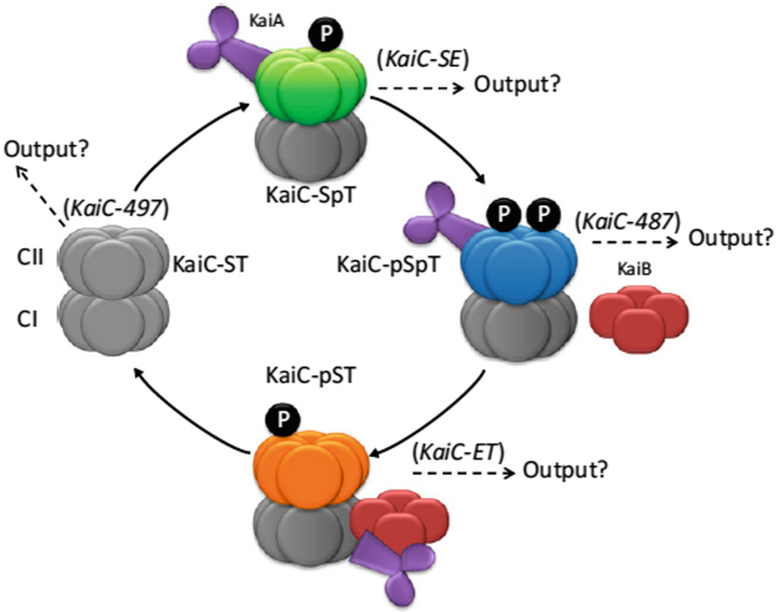


**Fig. 2.** Use of question marks to identify the question to be addressed in a research paper. Reprinted from Paddock et al. [17] with permission of the National Academy of Sciences, USA. (Color figure online)

A third use of question marks is to indicate lacunae in mechanism proposals. This occurs when researchers know enough to propose an overall operation, but are aware that it is a placeholder for more specific operations by parts not yet identified. For example, when Hardin et al. advanced the TTFL model, it was known that to inhibit transcription, a molecule must bind to DNA—but PER lacks a DNA binding region. This entailed that PER must interact with another molecule, which would then bind to DNA. The question marks at the top right in Fig. 1, thus, connote not only uncertainty about the target of the proposed feedback operation, but also the recognition that at least one unidentified molecule and operations involving it would need to be filled in.

A common way to identify such lacunae is to compare one mechanistic hypothesis to others that are, in at least one respect, better worked out. In their search for mechanisms, biologists often work comparatively between species. This is motivated in large part by the assumption that evolution is conservative—that genes, and hence the mechanisms for which they code, are inherited through phylogeny. Although extant species typically are not descendent from other currently existing species, they do have common ancestors from which they may have both inherited a common mechanism. Evolution, though, often introduces variability as another component that performs a very similar operation replaces an initial component. Even with this variability, researchers find it valuable to compare their current best account of what is taken to be the mechanism in one species with the account that has been developed for other species. Diagrams provide a particularly effective tool for this as they enable researchers to easily detect similarities and differences. A decade after Hardin et al.'s initial proposal of the TTFL model in the fruit fly, similar models were developed for several other model organisms. By this time additional parts and operations had been identified in the proposed mechanisms for one or another of these species, and this provided a motivation for comparing the proposed mechanisms to make inferences as to parts and operations that were still unknown in specific species. To this end, Harmer et al. [9] constructed Fig. 3 comparing what was then known about the circadian clocks in cyanobacteria (which at the time were still thought to employ a TTFL as their core circadian mechanism), fungi, fruit flies, and mammals.

To make it easier to compare the four proposed mechanisms, the authors used common shapes and coloring. Nodes shown in grey play the same role as PER in the fruit fly—inhibiting gene expression. A red, edge-ended arrow shows the target of inhibition. By this time PER (partnered with TIM in a PER:TIM dimer) was known to inhibit the dimer of CYCLE and CLOCK, which otherwise would promote the expression of *per* and *tim*. CYC:CLK is shown in yellow, as are the proteins thought to perform the same role in the other model organisms. In this context, Harmer et al. use questions marks to indicate unknown parts and operations. (In the case of operations, the question marks are located adjacent to dotted green lines.) For example, in the mouse clock the BMAL:CLK dimer was known to activate genes that transmitted the clock output to other biological activities. In the diagrams for the Neurospora and fruit fly clock, the comparable arrow is dotted and accompanied by a question mark. Likewise, whereas in the fruit fly mechanism a solid red inhibitory arrow is shown between the CLK:CYC dimer and the gene *clk*, in the mouse mechanism the comparable arrow between the BMAL:CLK dimer and the gene *bmal* is dotted and accompanied by a question mark. In cases where they were known, the proteins that activate gene expression at the promoter site are shown attached to the line preceding the symbol for gene expression. To indicate genes for which the activating protein is not known, the authors provide pink pentagons enclosing a question mark.

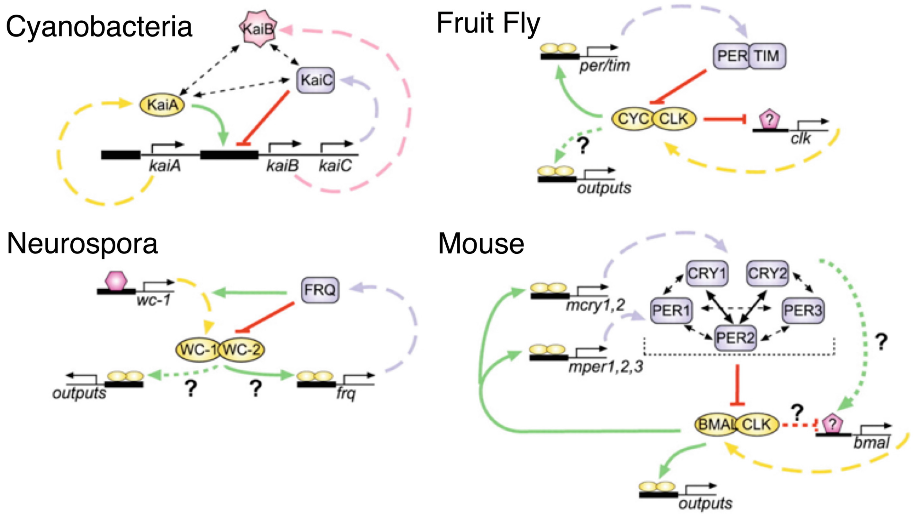


**Fig. 3.** Harmer, Panda, and Kay's comparison of the circadian clock mechanism in four model organisms. Modified with permission from the [9], by Annual Reviews, http://www.annualreviews.org. (Color figure online)

The three mechanism diagrams presented in this second section show variations on the common practice of using question marks to signify uncertainties or lacunae in the understanding of the mechanism. Figure 1 is the final figure of the research paper, and the question marks identify uncertainties about how, specifically, the proposed mechanism is supposed to work. Figure 2 served as the first figure in a paper that set out to answer the question posed by the question marks. Figure 3, from a review paper, inserts question marks into a diagram comparing the hypothesized mechanisms in four model organisms, in some cases in contexts in which an operation is known to occur in one organism's circadian mechanism, with question marks flagging where it might occur in homologous mechanisms in other organisms. Question marks thus reveal a way in which mechanism diagrams contribute to scientific reasoning.

## 3 Computationally Annotated Mechanism Diagrams

One major limitation of diagrams in providing mechanistic explanations is that they are static structures. Mechanisms, however, only produce phenomena if they are active—that is, if their parts are performing operations that change the state of other parts. As discussed above, one strategy for reasoning about such activity is to mentally animate a diagram—to imagine each part in turn carrying out its operation on other part(s), Hegarty [11] has not only described the practice but explored the limits of peoples' ability to mentally animate diagrams. These limits are obviously reached by the four diagrams in Fig. 3, due to their multiple operations that are non-linear or run in parallel with other operations. It becomes difficult to mentally animate such diagrams and even more difficult to ascertain whether the actual mechanism would behave in the way imagined. To overcome the limitations of mental animation, researchers often turn to computational models.

Computational modeling of the circadian mechanism began with the simplest pathway in Fig. 1, in which the protein X directly feeds back to the gene (identification of the dimerizing and mediating proteins in Fig. 3 came later). It can be mentally animated well enough to determine that it will produce oscillations, but not whether these oscillations will dampen vs. be sustained. Sustained oscillation is required to count as a mechanistic explanation of circadian rhythms. To demonstrate that this version of Hardin et al.'s TTFL could generate sustained oscillations, Goldbeter [7] constructed a computational model. As seen in Fig. 4, he linked the variables and parameters in its differential equations to the parts and operations in the TTFL mechanism. The positioning of the mathematical symbols in this annotated mechanism diagram served as what Jones and Wolkenhauer [12] refer to as a *locality aid* for constructing, modifying and reasoning about a computational model. (Note that some models include parameters that cannot be linked to any particular operation; these would be excluded from an annotated diagram and understood by other means.)

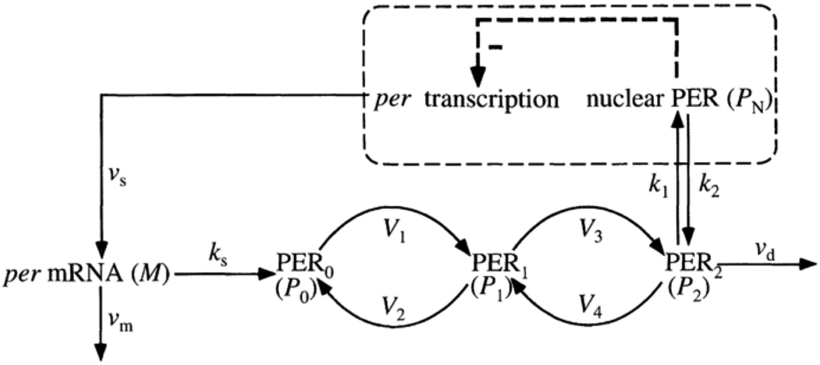


**Fig. 4.** Diagram of the PER TTFL annotated with names of the variables and parameters used in the computational model. Reprinted from [7], by permission of the Royal Society.

Figure 4 includes a few parts and operations not anticipated by Hardin et al., since by 1995 researchers had determined that after PER was synthesized in the cytoplasm, it had to be phosphorylated before it could be transported back to the nucleus. (PER$_0$, PER$_1$, and PER$_2$ indicate PER bound to zero, one, or two phosphates, and the nucleus is represented by the dashed box at the upper right.) After the name of each molecule, the variable tracking its changing concentration is shown in parentheses; for example, to the right of *per* mRNA is appended the variable (*M*). Next to each arrow denoting an operation is a parameter—three directly representing the rate of the corresponding operation ($v_s$, $k_1$ and $k_2$) and the others representing enzyme actions influencing rates (e.g., $k_s$ for the translation of mRNA into the protein PER). The computational model comprises five differential equations, one for each variable. The first equation, for example, obtains oscillations in the value of $M$ by subtracting a term that includes $v_m$ from a term that includes $P_N$ and $v_s$. The diagram makes clear the relevance of these terms to the value of $M$ and hence can help with the construction and mechanistic interpretation of that equation. Using the full five-equation model to run simulations, Goldbeter was able to demonstrate that, for a broad range of parameter values he characterized as biologically plausible, this TTFL would generate sustained oscillations.

In addition to assessing whether a proposed mechanism could produce the behavior researchers are trying to explain, computational models enable researchers to explore how the proposed mechanism would behave when perturbed in various ways. One reason researchers perform simulated experiments on computational models is that this enables them to better explain the behavior of the models. In this situation, the annotated mechanism diagram serves not just to help construct the equations used in the model but guides experiments that perturb it and their interpretation.

Such a use of computational models is illustrated in Fig. 5. In the two decades after Hardin et al. first proposed the TTFL model, researchers identified multiple feedback loops, some negative and some positive (one of these was already represented in the mouse circadian clock shown in Fig. 3 above). That generated the question of whether the whole complicated set of feedback loops is required to generate circadian oscillations or whether only some are essential. Relógio et al. [18] addressed this question by creating a computational model of the mammalian circadian clock. Like Goldbeter, they constructed an annotated mechanism diagram as a locality aid for the process of constructing the equations (Fig. 5). For each type of clock molecule that had been identified, the variable representing its concentration is appended to the geometric shape representing it. Likewise, each operation was associated with a relevant parameter in the model.

Once they established that the whole model would generate sustained circadian oscillations, they investigated whether the feedback loop involving PER and CRY or the loops involving BMAL, ROR, and REV-ERB could sustain oscillations on their own. The dashed line across the middle of the figure distinguishes these loops. To remove one or the other set of feedback loops from contributing to oscillations, the researchers fixed all the variables either above or below the line to a constant value (thus, preventing them from oscillating). When the variables above the line were set to a constant, the simulation ceased to produce oscillations, but when those below the line were so set, the simulation continued to produce oscillations. The diagram now supported the interpretation that only the feedback loops involving BMAL, ROR, and REV-ERB are needed for circadian behavior.

The practice of annotating mechanism diagrams with labels for variables and parameters facilitates their use in reasoning. The annotations guide the construction of computational simulations that allow researchers to go beyond the limits of mental animation in determining how the mechanism will behave. They also support designing and interpreting simulation experiments on the computational model.

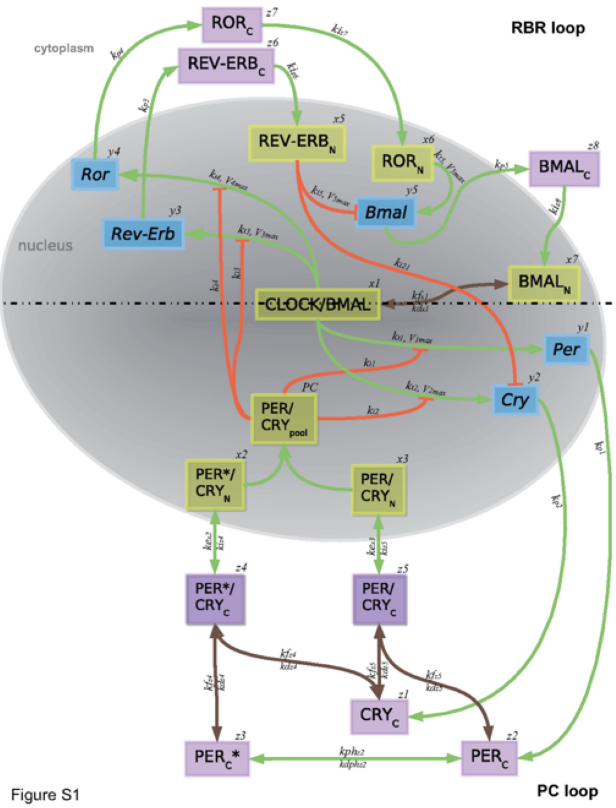


**Fig. 5.** Annotated mechanism diagram of the mammalian circadian clock developed by Relógio et al. as a basis for their computational model. Reprinted from [18]. (Color figure online)

## 4 Finding Patterns in Large Networks to Make Inferences About Mechanisms

Mechanism diagrams typically have been constructed by entering current knowledge of the parts and operations involved in producing a phenomenon of interest into a program offering graphics, such as PowerPoint. In the 21st century, however, systems biologists have offered an additional option: disregarding relevance to particular phenomena, they input very large datasets identifying entities and their interactions into a database from which specialized algorithms generate network representations. Further algorithms then can identify clusters that serve as candidate mechanisms. In a variety of contexts these network diagrams are enabling researchers to identify new mechanisms, discover new parts and operations of previously known mechanisms, and better understand how mechanisms are situated in a larger context.

To illustrate the practice, we start with an example from yeast systems biology. A widely used starting point for developing a network analysis is data specifying which proteins can form bonds with each other in a given species—an important capacity for implementing cellular functions. Techniques for determining which proteins form bonds with each other, such as the yeast two-hybrid technique, can be automated so as to test for hundreds or thousands of possible interactions. Data from studies testing for such protein-protein interactions (PPIs) are collected in large on-line databases such as the Biological Repository for Interaction Datasets (BioGRID). Other researchers can probe these datasets for use in their own studies and analyses.

To analyze PPI and similar data, researchers create network diagrams that represent genes or proteins as nodes and relations between them as edges. While mechanism diagrams are a type of network diagram, these networks are much larger and cannot easily be constructed using a standard graphics package. Accordingly, these network diagrams are commonly constructed using a specialized platform such as Cytoscape [19]. The database from which the network is constructed is represented as a table specifying nodes and edges as well as additional information, such as the expression level of each gene. A mapping function in Cytoscape translates the additional information in the tables into visible features such as the size and shape of nodes or the thickness of edges. A layout algorithm is then applied to situate the nodes in a two-dimensional array. Since a major goal of network analysis is to identify clusters of nodes that are highly interrelated, some variant of a spring-embedded layout procedure is often employed. Starting with either a random layout or the product of another layout algorithm, a spring-embedded algorithm functions like a spring, gradually reducing the distance between connected nodes that start far apart while pushing slightly apart those that are very close to each other. This reveals clusters of highly connected nodes, often referred to as *modules*. Researchers often can then relate these modules to more classically-identified mechanisms by drawing upon extant knowledge about where in the cell genes are expressed or what biological functions they perform. This knowledge can be accessed from other electronic resources, such as Gene Ontology [1]. Researchers can interpret modules whose components are not known to belong to any known mechanism as new candidate mechanisms. Also, for modules that do correspond to a known mechanism, they can infer that any nodes not corresponding to its known parts should be evaluated as candidate parts that may improve the mechanistic account.

To illustrate the strategy, Fig. 6 presents a network diagram of proteins involved in chromosome maintenance and duplication in yeast. Merico et al. [15] constructed the network using Cytoscape by importing protein-protein interaction data from BioGRID. Limiting their analysis to those proteins identified in Gene Ontology as located on chromosomes, they employed a spring-embedded algorithm to create a network diagram in which highly connected nodes were positioned near each other. Additionally, they incorporated information about the proteins' binding sites on chromosomes by mapping similar site locations to the color of nodes in the network representation (red: replication fork; green: nucleosome; blue: kinetochore; yellow: unknown chromosome components). In constructing the network they also incorporated data about how much the expression of each protein changes over the course of the cell cycle, mapping this onto the size of nodes. Finally, they used thickness of edges to represent the degree of correlation in the expression of pairs of genes.

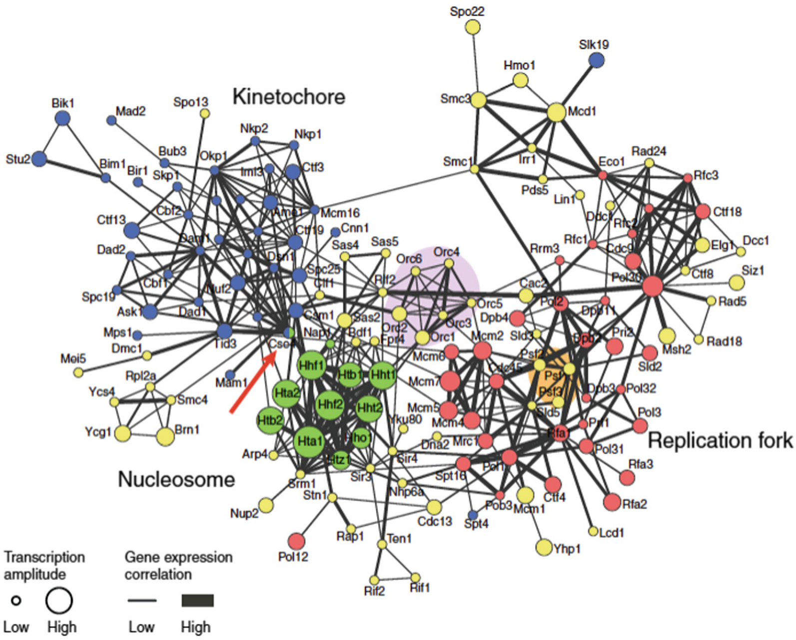


**Fig. 6.** A network representation of protein-protein interactions in yeast constructed in Cytoscape by Merico et al. to illustrates the strategy of guilt by association in inferring new parts and operations of a mechanism. Reprinted by permission from Nature/Springer/Palgrave, [15]. (Color figure online)

A few features are immediately observable. First, although information about location on the chromosome was not used to generate the layout (instead the layout was based on degree of protein-protein interaction), the blue, green, and red nodes that did reflect those locations form clusters. From the fact that these clusters of interacting nodes also bind nearby on the chromosome, the researchers interpreted them as corresponding to mechanisms. Second, the green nodes, representing proteins in the nucleosome, are especially large and connected by thick edges, indicating correlated dynamical change in the concentrations of these proteins. This further supports a mechanistic interpretation of these clusters.

While it is interesting that one can recover a mechanistic interpretation from the layout and feature mapping procedures of Cytoscape, what makes doing so worthwhile is the possibility of advancing new mechanistic hypotheses. To do this, researchers use a heuristic strategy known as *guilt by association*: whenever a node is not already annotated with a specific location or function, hypothesize that it has the same location or function as its neighbors. Nodes shown in yellow correspond to proteins that lacked a location specification. Three such proteins, Psf1, Psf2 and Psf3 (located in a region shaded in orange in Fig. 6), were positioned by the spring embedded algorithm among proteins belonging to the replication fork due to having multiple interactions with these proteins. Using guilt by association, the researchers inferred they played a role in chromosome replication. Although the information was not used in constructing the network, these proteins were known independently to be members of the GINS complex that is involved in the assembly of DNA replication machinery, indicating that guilt by association resulted in a correct inference. A similar line of reasoning leads to the hypothesis that Cse4 (indicated by a red arrow), which lies between the nucleosome and kinetochore clusters, figures in the interface of these two components. One can also use network analysis as a guide to new mechanisms when highly interconnected units lacking a specific annotation are found. In Fig. 6, Orc1-6, all shown as yellow nodes (in a region shaded in violet), are highly interconnected, suggesting they form a mechanism. Again, although the information was not used in constructing the network, they were known from other research to constitute the origin recognition complex (ORC).

By incorporating data about all known interactions, network analyses are able not only to recover knowledge of known biological mechanisms, but also to generate hypotheses about mechanisms not previously identified and about new parts and operations in existing mechanisms. But they can also serve another function—of situating mechanisms within larger contexts of interacting components within the cell. For the most part, mechanistic research looks inward to the component parts and operations of a mechanism. The resulting diagrams, such as those shown in Figs. 1, 2, 3, 4 and 5, present the mechanisms as self-contained, often with no inputs specified. In part this is an artifact of focusing on mechanisms responsible for circadian rhythms, which are generated endogenously by cells. Diagrams for other mechanisms (e.g., fermentation) may show an input and an output (e.g., glucose and alcohol), but they nonetheless treat mechanisms as largely autonomous from each other. Network diagrams like Fig. 6, while differentiating individual clusters, reveal numerous connections between them. This provides a way of identifying ways in which biological mechanisms affect each other beyond one mechanism simply providing an input to another. We return to a circadian example to illustrate this.

A variety of research has suggested that standard accounts of the circadian clock mechanism underestimated the range of genes/proteins involved in generating circadian rhythms, leading Zhang et al. [24] to use RNA interference techniques to individually knock down 17,631 known and 4837 predicted human genes in a U2OS human osteosarcoma cell-line. Using a luciferase reporter, they screened each for altered rhythms in individual cells. Among the thousands of genes that produced altered rhythms when knocked down, they selected 343 genes that yielded large and reliable alterations. One of their goals was to understand how the various genes identified in their study related to the core mechanism. However, the proteins coded by many of these genes did not directly interact with any components taken to be part of the core mechanism. To determine how they might be indirectly connected, the researchers drew upon the Entrez Gene and Prolexys protein-protein interaction databases to generate a protein-protein interaction network in which they identified the genes that in their screen altered rhythms. This yielded the network diagram in Fig. 7. The known clock genes are shown in light or dark blue as a densely interactive cluster in the center. The researchers drew upon existing mechanistic models of the clock to represent the edges between these genes as excitatory (green arrow) or inhibitory (red flat-edged arrow). Genes whose knockdown resulted in a short period are shown in green, those generating a long period in red, and those yielding high amplitude rhythms in purple. Shown in pink are genes that did not have a significant effect on the period or increase the amplitude of the rhythms but are intermediates between those that did and the core clock mechanism. The figure makes clear that in most cases one or two intermediates lie between the genes that did have an effect on the period or amplitude of circadian rhythms and the core clock genes. These genes need to be included in a mechanistic account of how knockdowns affect rhythms.

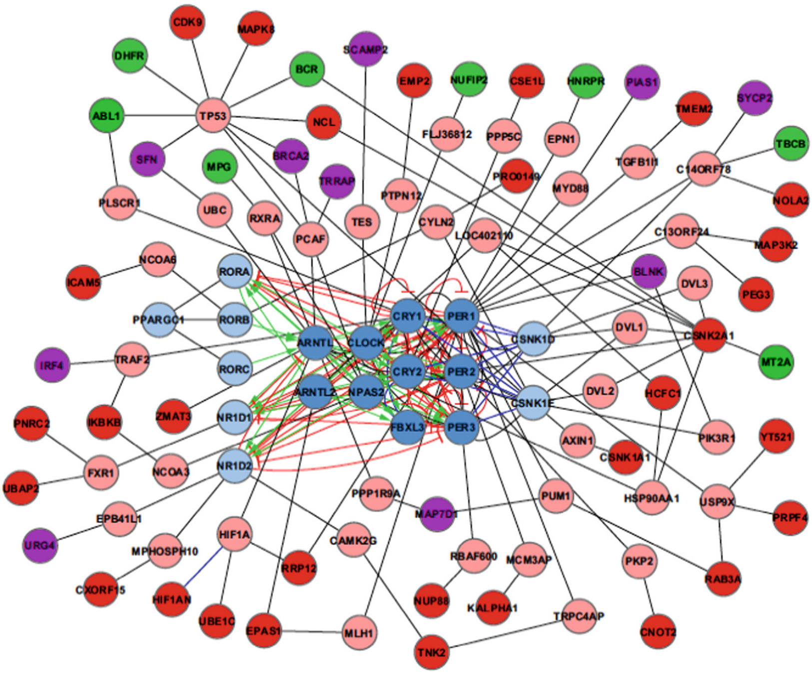


**Fig. 7.** Network diagram of genes that when knocked down result in altered periods or increased amplitudes of circadian rhythms. See text for details. Reprinted from [24], Fig. 5, © 2009 with permission from Elsevier. (Color figure online)

An interesting example of an intermediate gene is shown in the upper left. TP53, a major tumor suppressor gene, mediates between several proteins such as DHFR and CDK9 that the research showed to shorten or lengthen circadian periods. This suggests a role for TP53 itself in regulating circadian rhythms. The authors note that a study in the same year had identified BMAL1, a core clock gene, as a potential regulator of TP53. In Fig. 7 TP53 is only connected to BMAL1 (referred to in the figure as ARTNL) via another intermediate PCAF. The network analysis thus provides a guide for further investigation into how TP53 mediates the effects of knockdowns of other genes on the functioning of the core clock mechanism and whether this connects with its role as a cancer tumor suppressor.

## 5 Implications

One goal of this paper was to highlight the rich database of mechanistic diagrams on offer in the biological sciences, awaiting further attention from the diagrams research community as well as mechanistic philosophers of science. Our brief discussion of question marks, computational annotations, and network formats points to just a few of the neighborhoods in this database that could fruitfully be pursued in cognitive science experiments, ethnographic studies, case studies, and quantitative analyses. Although we focused on the molecular mechanisms responsible for circadian rhythms, similar diagram practices can be found in other areas of molecular biology and biology more generally, along with other innovations that arose within particular research communities and may then have spread.

Another important goal was to emphasize how diagrams support scientific reasoning. They play an underappreciated role in scientists' identification of new mechanisms, discovery of new parts and operations of already known mechanisms, and attempts to understand how a mechanism is situated in a broader environment that affects its functioning. Thus, diagrams are used not only to communicate proposed mechanistic explanations to an audience, but also to encourage reasoning about those explanations—leading at least to better understanding but often to tweaks or substantial changes as well. Diagram researchers who choose to pursue the research opportunities offered by diagrams of biological mechanisms can drill into diagram-aided reasoning, bringing to bear the methods of their own particular disciplines.

**Acknowledgement.** We gratefully acknowledge the support of NSF Grant 1127640.

## References

1. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., et al.: Gene ontology: tool for the unification of biology. Nat. Genet. **25**, 25–29 (2000)
2. Bechtel, W., Abrahamsen, A.: Explanation: a mechanist alternative. Stud. Hist. Philos. Biol. Biomed. Sci. **36**, 421–441 (2005)
3. Bechtel, W., Richardson, R.C.: Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research. MIT Press, Cambridge (1993/2010). 1993 edition published by Princeton University Press
4. Burnston, D.C., Sheredos, B., Abrahamsen, A., Bechtel, W.: Scientists' use of diagrams in developing mechanistic explanations: a case study from chronobiology. Pragmat. Cogn. **22**, 224–243 (2014)
5. Cheng, P.C.-H.: Probably good diagrams for learning: representational epistemic recodification of probability theory. Top. Cogn. Sci. **3**, 475–498 (2011)
6. Craver, C.F., Darden, L.: In Search of Mechanisms: Discoveries Across the Life Sciences. University of Chicago Press, Chicago (2013)
7. Goldbeter, A.: A model for circadian oscillations in the Drosophila period protein (PER). Proc. R. Soc. Lond. B: Biol. Sci. **261**, 319–324 (1995)
8. Hardin, P.E., Hall, J.C., Rosbash, M.: Feedback of the *Drosophila period* gene product on circadian cycling of its messenger RNA levels. Nature **343**, 536–540 (1990)

9. Harmer, S.L., Panda, S., Kay, S.A.: Molecular bases of circadian rhythms. Annu. Rev. Cell Dev. Biol. **17**, 215–253 (2001)
10. Hegarty, M.: Mental animation: inferring motion from static displays of mechanical systems. J. Exp. Psychol. Learn. Mem. Cogn. **18**, 1084–1102 (1992)
11. Hegarty, M.: Capacity limits in diagrammatic reasoning. In: Anderson, M., Cheng, P., Haarslev, V. (eds.) Diagrams 2000. LNCS (LNAI), vol. 1889, pp. 194–206. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-44590-0_19
12. Jones, N., Wolkenhauer, O.: Diagrams as locality aids for explanation and model construction in cell biology. Biol. Philos. **27**, 705–721 (2012)
13. Konopka, R.J., Benzer, S.: Clock mutants of Drosophila melanogaster. Proc. Natl. Acad. Sci. (USA) **89**, 2112–2116 (1971)
14. Machamer, P., Darden, L., Craver, C.F.: Thinking about mechanisms. Philos. Sci. **67**, 1–25 (2000)
15. Merico, D., Gfeller, D., Bader, G.D.: How to visually interpret biological data using networks. Nat. Biotechnol. **27**, 921–924 (2009)
16. Nersessian, N.J.: Creating Scientific Concepts. MIT Press, Cambridge (2008)
17. Paddock, M.L., Boyd, J.S., Adin, D.M., Golden, S.S.: Active output state of the Synechococcus Kai circadian oscillator. Proc. Natl. Acad. Sci. **110**, E3849–E3857 (2013)
18. Relógio, A., Westermark, P.O., Wallach, T., Schellenberg, K., Kramer, A., Herzel, H.: Tuning the mammalian circadian clock: Robust synergy of two loops. PLoS Comput. Biol. **7**, e1002309 (2011)
19. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., et al.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. **13**, 2498–2504 (2003)
20. Sheredos, B., Bechtel, W.: Sketching biological phenomena and mechanisms. Top. Cogn. Sci. **9**, 970–985 (2017)
21. Stull, A.T., Hegarty, M.: Model manipulation and learning: fostering representational competence with virtual and concrete models. J. Educ. Psychol. **108**, 509–527 (2016)
22. Tversky, B.: Semantics, syntax, and pragmatics of graphics. In: Holmqvist, K., Ericsson, Y. (eds.) Language and Visualization, pp. 141–158. Lund University Press, Lund (2004)
23. Tversky, B.: Visualizing thought. Top. Cogn. Sci. **3**, 499–535 (2011)
24. Zhang, E.E., Liu, A.C., Hirota, T., Miraglia, L.J., Welch, G., Pongsawakul, P.Y., et al.: A genome-wide RNAi screen for modifiers of the circadian clock in human cells. Cell **139**, 199–210 (2009)

# A Survey and Evaluation of Diagrams for Navya-Nyāya

Jim Burton[1(✉)], Lopamudra Choudhury[2], and Mihir Chakraborty[2]

[1] University of Brighton, Brighton, UK
j.burton@brighton.ac.uk
[2] Jadavpur University, Kolkata, India
choudhuryl@yahoo.com, mihirc4@gmail.com

**Abstract.** Navya-Nyāya, "The New Reasoning", is a formal philosophical logic developed in India from the 11th to the 17th centuries CE, and which builds on the older traditions of Nyāya and Vaiśeṣika. Not surprisingly, Navya-Nyāya is fundamentally different from classical Western logic and from the meanings ascribed to traditional logical diagrams. For instance, although it is not entirely correct to describe Navya-Nyāya as extensional or intensional, it has an intensional flavour: abstractions are built up from concrete individuals of which we know only their possession, or not, of certain properties. In this paper we look at the implications of these semantics for the use of logical diagrams in Navya-Nyāya. We survey the use of diagrams in modern studies of Navya-Nyāya, notable examples having been produced by Wada, Das and Ganeri. We use notions of *well-matchedness*, *iconicity* and Cheng's recent framework to analyse the effectiveness of the notations in the context of their intended purposes.

## 1 Introduction

Indian logic is a collection of texts, traditions and techniques stretching back at least 2000 years. The first major text of the Nyāya school is Gautama's Nyāya-sutra, circa 200 CE, which presents a detailed, closely argued but nevertheless informal and analogical style of reasoning. The development of Nyāya into something more precise took place beginning around the 11th century CE with the writings of Udayana, and by the time of Gaṅgeśa's work *Tattvacintāmani* in the 14th century it was appropriate to describe the developments as a new school, the Navya-Nyāya. Achievements of the new school include a system for constructing unambiguous expressions imposed on the abundantly ambiguous Sanskrit language, a detailed analysis of logical relations, a system akin to quantification and a foundational theory of number [7, p. 91].

None of the original Sanskrit authors used diagrams. The long tradition of logical diagrams in the West [16] is unseen in India, for reasons that are unclear; we speculate it may be partly related to the *śruti* tradition which privileges the conveying and acquiring of knowledge orally, not visually. Since Navya-Nyāya is a linguistic activity whose main tool is an augmented Sanskrit grammar, the

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 280–295, 2018.
https://doi.org/10.1007/978-3-319-91376-6_27

early absence of diagrams may be expected. A number of modern authors have used a variety of diagrams, however, and this paper is a survey and analysis of their efforts.

Our main form of analysis is to consider how fit for purpose the notations used are from the perspective of *well-matchedness* or, equivalently (for our purposes at least), how the notation makes use of *iconicity*. We also apply Peter Cheng's recent framework [2] by way of contrast and as a modern, cognitively-focused approach.

In the course of this study we will explain our view that spatial diagrams (Euler, Venn, linear diagrams etc.) are poorly matched to the meaning of Navya-Nyāya. The semantics of Navya-Nyāya means that intension, i.e. individuals and their properties, is of the first importance. Thus, rather than notations designed to show subsumption, intersection, disjointness and so on, diagrams based on various kinds of network are most appropriate. Of these, we examine the diagrams of Chi [3], Wada [24], Das [4] and Ganeri [7]. Each of these authors had different (frequently overlapping) purposes in mind when developing their notations, and so we approach the analysis in ways sensitive to the differing contexts. The existing notations are intended for exposition, not for use as diagrammatic logics. By our criteria (set out in Sect. 2) none of the notations is currently well adapted for reasoning, i.e. for visualising inference and the development of diagrammatic inference rules that are distinctively Nyayān.

In Sect. 2 we summarise details of the frameworks by which we analyse the notations, whilst in Sect. 3 we explain some basic terms and concepts from Navya-Nyāya. In Sect. 4 we survey the extant diagrammatic notations of Navya-Nyāya, with a commentary on their fitness-for-purpose given via the methods outlined in Sect. 2. We conclude by analysing the requirements of an analytical notation that will highlight the key features of reasoning in Navya-Nyāya in ways that enable and support inference.

## 2 Methodology

For Charles Peirce, diagrams were the preferred vehicle for reasoning. He believed that diagrams allow us to see the relevant relations at work in a logical expression and to manipulate those relations directly. The earlier version of Peirce's sign classification divides the modes in which signs convey meaning in three types: *Icons*, which depict by resemblance, *Indices*, which depict by directly "pointing to" or indicating something, and *Symbols*, which depict by convention. Each modality has a type of information it is best suited to convey: icons should be used for theorems, indices for existential statements, and symbols for general laws [20]. Of the sign modalities, iconicity is the more difficult to grasp, since it depends on the problematic idea of resemblance [22]. This resemblance is intended to be structural or relational, rather than pictorial. Under this interpretation a sign resembles its object if and only if study of the sign can yield new information about the object [12]. A diagram is a special kind of icon whose parts stand in the salient relations to each other, and which comes with rules that allow us to manipulate the parts in reliably valid, illuminating ways. This aspect of iconicity is also captured under the name *well-matchedness* [10]. A well-matched notation shares certain important characteristics with the domain it depicts. For instance, Euler diagrams are well-matched to propositions and assertions about sets because they have "the potential to directly capture pertinent aspects of the represented artifact" [11]. Well-matchedness is sometimes discussed as depending on the existence of a structure-preserving map (homomorphism) between syntax and semantics; this is a necessary but not sufficient condition. Ambrosio emphasises the active, contextual nature of iconicity: "'Constructing' an icon amounts to discovering, and selecting, relevant respects in which a representation captures salient features of the object it stands for." [1] Thus, designing and using iconic diagrams sheds light on the underlying structures of reasoning and on what it means to think diagrammatically.

To provide contrast we will also apply concepts from the framework proposed by Cheng in *What Constitutes an Effective Representation* [2]. This work approaches the questions outlined above both heuristically and in ways supported by results from cognitive science. Cheng identifies 19 criteria which are organised in two main categories: *Access to Concepts*, for criteria which relate to the ease of translating from external to internal representations, and *Generating External Representations*, which collects criteria relating to building new content in a notation. We will focus on the first category, *Access to Concepts*, in which, although using different terminology, several of Cheng's criteria reflect a similar perspective on notational effectiveness to that of well-matchedness and iconicity and seem appropriate for our needs. These include A1.1, *One token for each type*, A1.2, *Reflects structure of concepts*, A3.2, *Coherent encoding of primary concepts* and A3.3 *Overarching interpretive scheme*. In A1.5, *Iconic expressions*, "iconic" is used to mean "clearly recognisable and particularly memorable" [2], which is quite different to Peirce's usage. For reasons of space we are not able to consider each criteria against each notation, and choose the criteria which highlight the most relevant issues in each case.

## 3 The Basic Terms and Concepts of Navya-Nyāya

Navya-Nyāya is a complex and detailed system of which we will only attempt to describe the aspects needed in order to understand the diagrams in Sect. 4. Useful introductions can be found in Ganeri [5] and Guha [9]. Like all philosophical logic, Navya-Nyāya is concerned with the analysis of thought. It combines techniques of rhetoric (how to construct a convincing argument), epistemology (the analysis of truth and knowledge) and logic proper (how to draw valid conclusions on the basis of existing evidence), three fields that have been considered distinct from each other in the Western tradition. Nyāya (in its original and "new" forms) is closely linked with the Vaiśeṣika school, from which it takes their seven-part ontology wholesale. The Vaiśeṣika ontology divides objects in the real world into seven types. A positive entity is either a *universal*, a *quality*, a *motion*, or a *substance* (which may be *compound* or *atomic*) or an *individuator*. The types are distinguished from each other by inherence (*samāvaya*): the number of entities each type inheres in and the number which may inhere in it. Nothing may inhere in a universal, while universals inhere in qualities. For example, the universal *blue* inheres in a quality, *blueness*, which is a particular shade of the colour. Qualities inhere in substances; blueness may inhere in a pot, for example. A pot is a compound substance composed of smaller parts. The pot inheres in each of those parts, all the way down to the atomic substances it is made of, which inhere in nothing. The other type of entity which inheres in atomic substances is the individuator, which allows us to distinguish one atom from another. The seventh type of entity is the negative entity of *absence* (*abhāva*). Since the Nyayāikas believe that every cognition has some content, when we perceive the absence of a pot on the ground, that absence is an entity with real existence. What we perceive in this case is an *abhāva* which is counter-positive (*pratiyogi*) to pot-ness (the quality of being a pot) and which has an absential-spatial location in the ground. This is an example of *relational* absence, and there are other varieties dealing with temporal absence and inequality. Nothing inheres in a negative entity. As we will see in Sect. 4, this model of reality is reflected to a greater or lesser degree in the diagrams used to visualise Navya-Nyāya.

The usual introduction to Navya-Nyāya is via the *anumāna*, or inferential schema. There are two varieties, with three and five steps respectively. The example below is the five part variety, or *pararthanumāna*, taken from Ingalls [13]:

**Anumāna 1.**

1. **Thesis:** Word is non-eternal.
2. **Reason:** Because it possesses the property of being produced.
3. **Statement of pervasion and example(s):** What possesses the property of being produced is seen to be non-eternal, as a pot. What possesses the property of not being produced is seen to be eternal, as the soul.
4. **Application:** It (word) is like this (i.e. possesses the property of being produced).
5. **Conclusion:** Therefore word is non-eternal.

When this logical structure was introduced to the West in 1824 by Colebrooke [6], it was named the "Indian syllogism". This choice of name formed part of the misunderstanding of Indian logic in the West that was to last the best part of a century. Compared to Aristotle's syllogistic, *pararthanumāna* is inadequate in several respects, particularly its repetition (steps 1 and 5) and the redundant and distracting appeal to examples in step 3. The history of these misunderstandings is described by Ganeri [6].

In fact, *pararthanumāna* bears no real resemblance to syllogistic reasoning. It is not concerned with classes of things, relations between classes or membership of those classes. Its content concerns individuals, their "pervasion" by properties and an inference that can be made thereby. Mullick explains that the distinction between implication and entailment is key to (mis)understanding the nature of inference in *pararthanumāna*, which should be seen as an inference schema or rule than a series of propositions. Implication "holds by virtue of the meaning-content of propositions rather than their truth-values, [and] this must itself be because of the concepts they contain." Mullick described the inference taking place in *pararthanumāna* as "conceptual implication". Furthermore, *parathanumāna* is a schema because although each example discusses particular objects (words, hillsides and so on) the role of these "paradigmatic" objects is as placeholders for any objects or properties that stand in the given relations to each other [17].

Two of the key notions at work in anumāna are *locus* and *locatee*. In the example above, the loci are *word* (which is the *pakśa*, or locus that we want to reach a conclusion about), *pot* (the *sapakśa*, or example which is claimed to be like the *pakśa*) and *soul* (the *vipakśa*, the example which is claimed to be unlike the *pakśa*). The locatees are the properties of being eternal, of being produced, and the absence of each of those properties. Non-eternal-ness is the *sādhya*, the target property or thing we want to prove is true of the *pakśa*.

The treatment of relation is strongly intensional. The eternal-ness of an entity (such as the soul) is conceived of as a quality which inheres in a locus. The quality is thus delimited by and particular to the locus (through a process called *avacchedakatva*, the delimitor/delimited relation). The eternal-ness of a given soul is not universal eternal-ness, and is not equivalent to the eternal-ness of a separate entity. This approach extends to every cognition. To express the notion "a pot is on the ground" the Nyayāikas construct a statement equivalent to "contact delimited by pot-ness inheres in the ground". This is their means of avoiding the ambiguity of ordinary speech.

One of the most distinctive features of Navya-Nyāya is the treatment of absence (*abhāva*) and negation. Recalling the example of our perception of the absence of a pot on the ground, that absence is an entity with real existence. In this case, "absence delimited by pot-ness inheres in the ground". Returning to the *pararthanumāna* example above, for word to be non-eternal means that the absence of eternal-ness is located in word; this absence is also located in pot but is not located in soul.

Our second example of *pararthanumāna* is the most commonly cited, which we will use to explain the third step and use of examples:

**Anumāna 2.**

1. **Thesis:** This hill is fiery.
2. **Reason:** Because it is smoky.
3. **Statement of pervasion and example(s):** What possesses the property of being smoky is also fiery, as in the kitchen hearth. What possesses the property of not being smoky is not fiery, as in the lake.
4. **Application:** This hill is so.
5. **Conclusion:** Therefore this hill is fiery.

The third step has been taken by various authors as a predicate logic expression such as $\forall x.S(x) \rightarrow F(x)$. This fails to reflect the intended meaning in several ways. Apart from $S(x)$ and $F(x)$ being an inadequate way to represent the notions of locus, locatee and *avacchedaka*, this is a statement about pervasion, as the name suggests. Entities which are pervaded by smoke are necessarily pervaded by fire. This is about the relation between the *sādhya* (target) and *hetu* (reason) properties. For the *hetu* to be a reliable reason property, the *sādhya* must be seen wherever the *hetu* is seen. It is quite possible that the *sādhya* is seen in some instances where the *hetu* is not; a red-hot iron is said by the Nyayāikas to possess fire but not smoke. But it must certainly be true that the extent of the *hetu* is strictly contained within that of the *sādhya*. This leads to the Buddhist logician Diṅnāga's "reason with three characteristics" [7, p. 115]:

> [A proper reason must be] present in the site of inference and what is like it and absent in what is not.

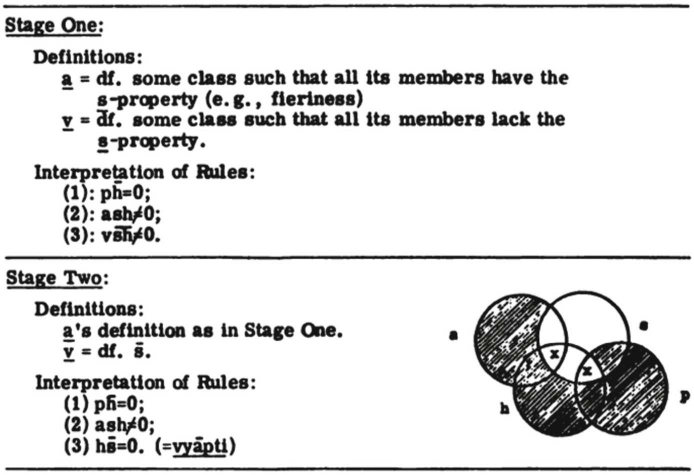
How do we know the examples are well-chosen, and what is "like" and "unlike" the site of inference? How do we know that there are no counterexamples which are yet to be observed (a Black Swan event)? As pointed out by Diṅnāga, we are concerned here with *like* and *unlike* in respect to the *sādhya*, not the *hetu*. In commentary on Diṅnāga's work, Dharmakirti went on to claim that there are three ways in which *hetu* and *sādhya* may be linked: both properties may be linked to (possibly caused by) the same phenomenon, may be linked by metaphysical causation, or we may reach a conclusion about their interdependence based on the lack of counterexamples. After observing some small number of examples we can be satisfied, and stop looking. [7, p. 121] The reliance on the examples in *pararthanumāna* can be problematic for readers hoping to understand Navya-Nyāya, leading some to think that this is an informal case-based reasoning by analogy. However, if we accept the validity of the existence of likeness and unlikeness classes and accept the assumption that we are able to identify an example from each, then the status of *pararthanumāna* as a formal inference schema is clear. Matilal puts it thus ([14], quoted in [19]):

> In short, the Nyāya strategy is to appeal to our intuitions about knowledge, in order to learn something about reasoning and not vice versa.

The last concept we need to mention is *paryāpti sambhanda*, or the "relation of completeness" [9, p. 50]. This phenomenon occurs when two or more entities are joined in a grouping relation, such as the "two-ness" that inheres in each element of a pair, $(x, y)$. Two-ness inheres in each of $x$ and $y$ but only when each is considered as part of a pair. If we consider either element in isolation, two-ness is no longer observed. *Paryāpti* plays a key role in the Nyayān analysis of cognition, but is not always given special treatment in diagrams. Next, we describe the diagrams of Navya-Nyāya and analyse their effectiveness.

## 4 The Diagrams of Navya-Nyāya

The earliest use of something approaching a diagram in a text on Navya-Nyāya appears to be from Goekoop in a book published in 1967 [8]. However, although the figures in Goekoop's book help the exposition, they are informal and have no analytic content. From a Peircean point of view these are images rather than diagrams, since the information provided is exactly that which was needed to construct it. A more diagrammatic style is used by Chi in his 1969 book *Buddhist Logic* [3]. Chi uses Euler diagrams with shading and inscribes an "x" to indicate the non-emptiness of regions, as seen in Fig. 1.



**Stage One:**

**Definitions:**

a = df. some class such that all its members have the s-property (e.g., fieriness)

v = df. some class such that all its members lack the s-property.

**Interpretation of Rules:**

(1): $p\bar{h}=0$;

(2): $ash\neq0$;

(3): $v\bar{s}\bar{h}\neq0$.

**Stage Two:**

**Definitions:**

a's definition as in Stage One.

v = df. $\bar{s}$.

**Interpretation of Rules:**

(1) $p\bar{h}=0$;

(2) $ash\neq0$;

(3) $h\bar{s}=0$. (=vyāpti)

**Fig. 1.** Chi's diagrams [3]

The major problem is the misplaced use of circles to represent such notions as "some class such that all its members have the s-property". This statement does have some meaning in Navya-Nyāya, and could be explained as *anugama*, which is the technique of "including diverse objects of the same category and sometimes even of different categories into one group by a single statement" [9, p. 209]. But *anugama* plays no role in this inference which is essentially intensional. The s-property is limited, in each case, by its location in an individual. The s-property of one is not the s-property of another. From the text it certainly appears that Chi is aware of this and does not suggest that *anugama* is involved, but uses a set-theoretic interpretation and spatial diagrams as a common ground to compare several Indian logics. Nevertheless, this usage violates well-matchedness. Spatial diagrams, which are widely agreed to be a very good match for logic in the western tradition (i.e. they are iconic, in that context) [15], will not help us to understand anumāna. If the benefits of spatial diagrams for Western logics reside in the fact that they can represent classes and relations "as they really are", this is not the case for Navya-Nyāya. Schayer had formed this view in the 1930s: "In the Indian example [of the 'syllogism'] there is no such thing as the relation of subsumption and the attempt to represent the inference with the help of Euler's circles has to be rejected." [18] In Cheng's terms, this representation fails to satisfy A1.2, *Reflects structure of concepts* and A3.2, *Coherent encoding of primary concepts*, among other criteria.

A more promising approach using node-link diagrams is found in Wada's work [23,24]. In the introduction of [24] he states that the diagrams are intended to illustrate *vyapti* (the connection between the fact that needs to be proved and our reason for believing it), not for reasoning with and not to explain the Vaiśeṣika ontology. Nodes represent qualities and substances and edges the relations between them, making this notation inherently more well-matched than any use of spatial diagrams. In Fig. 2, three types of edge are used: a plain edge (—) to represent inherence (*samāvaya*), an arrow (→) for the delimitor relation (*avacchedakatva*), and a double arrow (⇒) for *nirūpakatva*. *Nirūpakatva*, the determinor/determined relation, is a type of *avacchedakatva* that determines a relation or entity in a given context. Wada explains the example in Fig. 2 as follows. A pot-maker may be the cause of certain pots and a cognition of this fact would include the cognition of the pot-maker as qualified by cause-ness. However, the same individual may be the cause of multiple things (such as children, perhaps). If the most salient thing which he causes is the pots that are produced, then the cause-ness in question is itself qualified by an effect-ness which is qualified by pot-ness [23]. This is illustrated in Fig. 2 by the ⇒ edge.

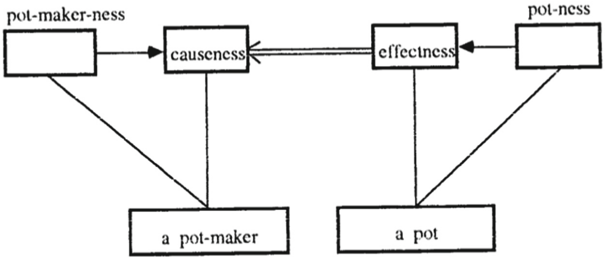


**Fig. 2.** A node-link diagram from [24]

Wada's diagrams admit a fourth kind of edge, a dashed line which depicts "any relation whose existence is negated". In Fig. 3 the dashed line denies any relation between the colour red and the kitchen.

Overall, Wada's diagrams are a much better match for Navya-Nyāya than any use of spatial diagrams could be. Most importantly, the node-link structure reflects intension. The resulting structure comes close to reflecting the authentic ontology of a cognition. Problems include the dashed edge and treatment of absence; to what entity does the dashed line relate? The inherence edge doesn't indicate the direction of the relationship. A more serious problem is that there may be non-unique readings of diagrams with several *nirūpakatva* edges (e.g. an identical diagram for the distinct cognitions arising from "man with stick in hand" and "stick in hand of man"). Ramesh Chandra Das provided a solution to this problem, which is explained below. More generally, a relatively small fragment of Navya-Nyāya is covered. For example, how should we express diagrammatically the different kinds of absence? How should we depict relations

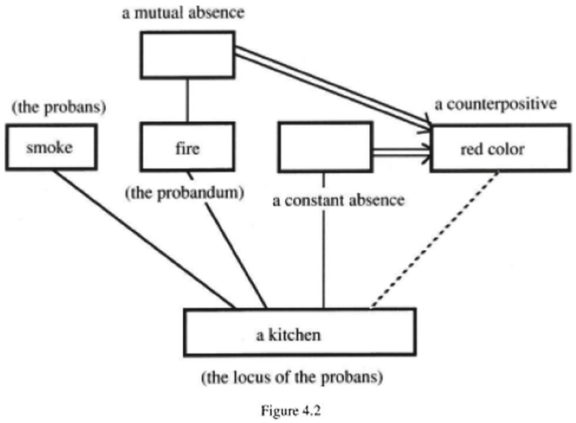


**Fig. 3.** Abhāva in Wada's diagrams [24]

that involve completeness (*paryāpti*)? Doing so may require a multiplying profusion of edges, nodes and labels that would obscure the underlying meaning. For instance, representation of completeness (*paryāpti*) should immediately highlight its "grouping" nature, but there is no device in Wada's diagrams that could do this. In this respect, the diagrams could be considered to fail Cheng's criteria A3.1, *A format for each class of primary concepts*. The alternative view is that Wada does not take completeness or variations in absence to be crucial for his needs, but both concepts can play a crucial role in inference. Other of Cheng's criteria highlight ways in which Wada's diagrams could be improved: the dashed edge does not faithfully reflect conceptual structure (A1.2).

Das introduced a form of extended node-link diagram in his 2006 work [4]. His diagrams are similar to Wada's but richer syntax makes more types of edge available, exposing more of the underlying semantics. Figure 4 shows the cognition "[the] pot is not on [the] ground", and does so more fully than is possible using Wada's notation.

The outer rectangle drawn around the inner network of nodes and edges represents *cognition*, providing the opportunity to juxtapose separate cognitions, although none of Das' examples do this. Edges can be drawn sourced on other edges, not just entities, enabling the representation of the (legitimate) state of affairs when an entity is qualified by a relation, rather than a quality or attribute. Figure 5 shows the components available in Das' notation.

In an important contribution to the formality of the diagrams, Das demonstrates how to reconstruct a Navya-Nyāya expression from a diagram. This relies on the addition of the meta-diagrammatic feature of a numbered circle to label *contextual properties* (Das refers to it as "an artificial device"). Contextual properties are those which determine other entities, i.e. those which are adjuncts in the *nirūpakatva* relation. The order in which contextual properties are described affects the meaning of the expression, similar to the order of quantifiers in

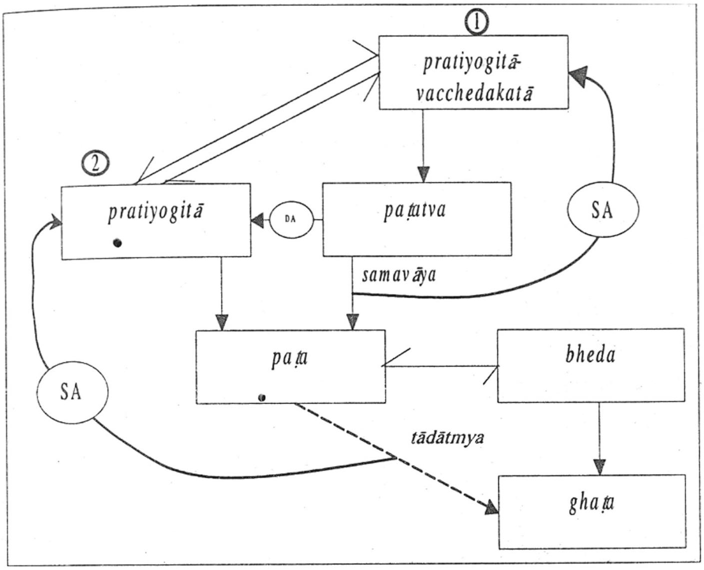


**Fig. 4.** Das' extended node-link diagrams [4]

predicate logic. The numbered circles in Das' diagrams that label contextual properties are read in order, producing the desired meaning by reimposing the linearity of the Sanskrit expression. The absence of this feature means Wada's notation is subject to ambiguity.

The orientation of nodes and edges has semantic import in Das' diagrams; a set of nodes and edges arranged horizontally has a different meaning to the same nodes and edges in a vertical arrangement. In Fig. 6, because the nodes in the cognition on the left are arranged vertically, we read the arrow as the locus/locatee or property/posessor relation. In the right-hand cognition, the arrow could be any relation and (in the absence of a label) all we know is that $X$ is related to $Y$.

A label is placed in the circle of the delimitor (*avacchedaka*) edge to show the kind of delimitor that is in effect. **DA** stands for *dharam avacchedaka* (attributive delimitor). Other possible delimitors are **SA** (*sambhanda avacchedaka* – relational), **KA** (*kala* – temporal) and **AA** (*aṁsa avacchedaka* – spatial). This vocabulary is enough for Das to give a detailed, unambiguous depiction of the cognition "the pot is on the ground" (*ghatavad bhūtalaṁ*), as seen in Fig. 7.

As well as being more formal, the notation developed by Das is more comprehensive than that of Wada in that it corresponds to a larger fragment of Navya-Nyāya. However, the notation relies on a larger number of conventional features, detracting from its iconicity. These features are warranted since they express precise meaning, but don't help us select the essential characteristics of inference. To understand how we can manipulate a internal conceptual model formed from a reading of these diagrams we need to internalise a lot of rules, some of them (e.g. orientation) quite arbitrary. Cheng's criteria make it clear

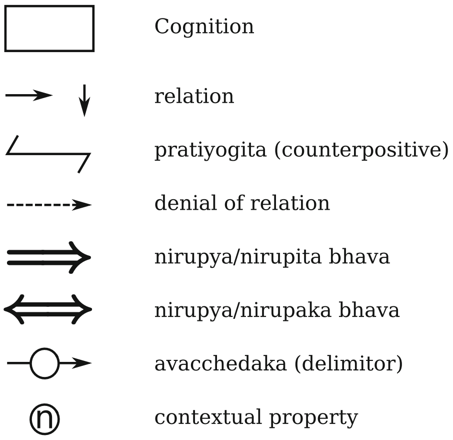


**Fig. 5.** The syntax of Das' diagrams

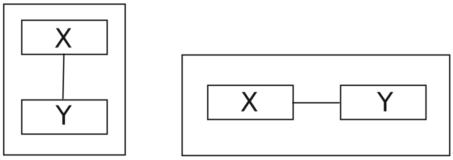


**Fig. 6.** The semantics of orientation in Das' diagrams

that Wada's goal of providing a comprehensive picture of cognition is double edged. It means that the notation performs highly in reflecting the structure of concepts (A1.2), directly depicting the structure of cases (A1.3) and, especially, invoking more structured internal representations (A2.3). Contextual properties enable the notation to provide an reliable overarching interpretive scheme (A3.3). These benefits come at the cost of complexity. Consider the use of the plain arrow ($\longrightarrow$) to denote arbitrary relations; it would not be possible to provide a distinct, meaningful token for each type (A1.1, A1.5) in this case but the goal of proving a comprehensive picture means the same token is used for relations with very diverse meanings, leaving the reader to "fall back" on the labels and violating A2.1, *Prefer low cost forms of processing.*

Ganeri uses diagrams extensively in his 2001 book *Philosophy in Classical India* [7]. The style of the diagrams arises from Ganeri's observation that the $11^{\text{th}}$ century scholar Udayana's justification of the Vaiśeṣika ontology differentiates between entities on the basis of inherence (*samāvaya*), and that these differences can be seen as properties of a directed graph. Types in the ontology are distinguished by the number of entities they can inhere in, and the number of entities that can inhere in them. Thus, what distinguishes a universal from the graph-theoretic point of view is that it is a node with incoming valency of 0. A universal must be "exampled" in at least two qualities, so its outgoing valency must be at least 2. Qualities and motions are inhered in by universals, so their incoming valency is 1. Substances are inhered in by entities which are themselves inhered in (i.e. qualities and motions), so they can be distinguished by their depth in the graph (the length of the paths that reach them) which must be at least 2. Compound substances inhere in their parts (which explains why some nodes representing substances may have a depth greater than 2), and atomic substances inhere in nothing, so their outgoing valency is zero. Individuators are unique in having an incoming valency of 0 (like universals) but an outgoing valency of 1.

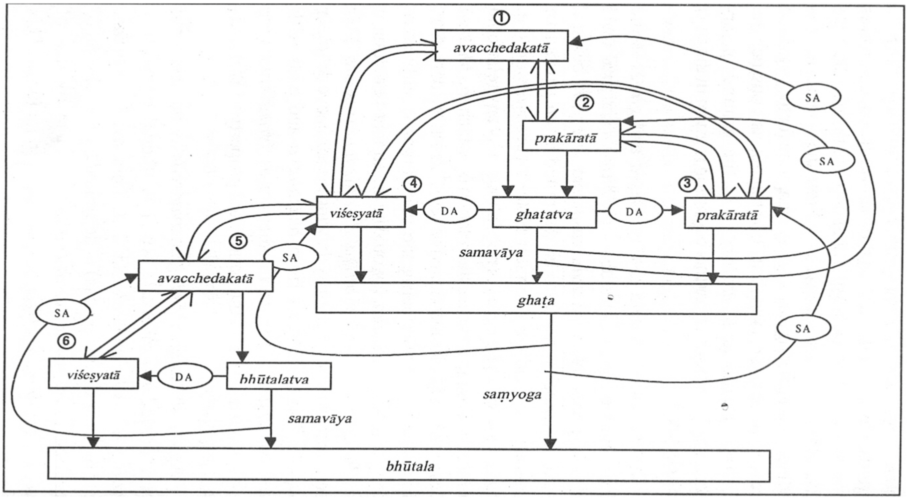


**Fig. 7.** Das: "[the] pot is on [the] ground" [4]

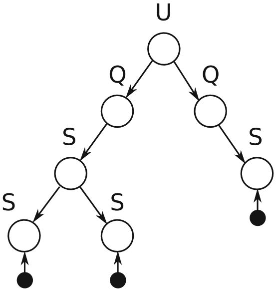


**Fig. 8.** An ontology in Ganeri's graphs

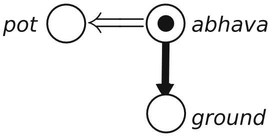


**Fig. 9.** Absence in Ganeri's graphs

Each node in Fig. 8 is a *bhāva* (positive entity) and an arrow between nodes $n_1$ and $n_2$ means that $n_1$ inheres in $n_2$. The set of edges is the extent of the inherence relation in the depicted cognition. Figure 8 shows an ontology in which one universal inheres in two qualities which inhere, respectively, in a compound and atomic substance, and so on. The solid black nodes represent individuators. After explaining that individuators were considered redundant by Ragunātha [7, p. 77] Ganeri leaves them out of subsequent graphs.

To represent absence (*abhāva*) a new type of node and two new edges are introduced, as in Fig. 9. The $\Longrightarrow$ edge denotes counterpositiveness (*pratiyogita*) while the heavy edge is the absential spatial relation. Unlike Wada and Das, no non-existent "denial of relation" is depicted. The approach to absence allows a simple illustration of the dielethic aspect of Nyāya cognition, whereby a proposition and its negation can be true at the same time. This is an *unpervaded occurrence* [7, p. 88], an example of which might be a jug with a red body and blue handle. The composite jug is inhered in by the colour blue but the absence of blue is located in it also. See Fig. 10.

A branching edge is used to represent completeness (*paryāpti*). Figure 11 illustrates the situation where the "quality 'two' inheres in both members of a pair of substances, another quality 'two' inheres in another such pair, and all the qualities 'two' have inhering in them a single universal 'two-hood'" [7, p. 91].

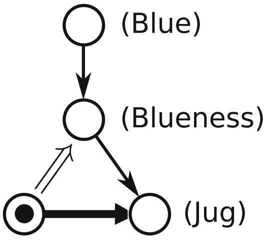


**Fig. 10.** Unpervaded occurence

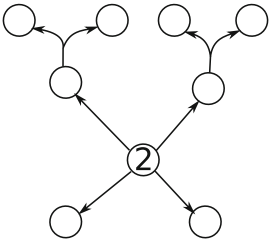


**Fig. 11.** Completeness in Ganeri's graphs

Ganeri's graphs are ideally well-matched to Udayana's explanation of the Navya-Vaiśeṣika ontology, which *is* a directed graph. The treatment of completeness is iconic in that it has the relation (grouping, or joining) that it represents. The depiction of absence includes no redundancy. The well-matchedness allows for a dialogue between graph-theoretic properties of a diagram and the underlying model. For instance, one Nyayān controversy centred on whether qualities and motions are really distinct entities, or whether a motion is actually a form of quality; in the graph-theoretic model their equivalent valencies mean they are indistinguishable, reflecting the simplifying revision argued for by later Nyayāikas [7, p. 77]. Overall, the notation is minimal and elegant relative to Wada and Das, but it manages this because it is more specialised to the ontology and covers only a small fragment of other concerns (the processes of cognition and inference). Within the context of its purpose, the graphs performs well on Cheng's framework, providing one token for each type (A1.1), and reflecting the structure of concepts and the structure of cases more accurately than other notations (A1.2, A1.4). The structure of directed graphs makes use of a low-cost form of processing (A2.1). When we look outside the target domain of the Navya-Vaiśeṣika ontology, certain concepts (such as the determinor relation, see below) are missing (A3.1, A3.2).

We now consider the task of using Ganeri's graphs to depict the inference embodied in anumāna. Figure 12 applies the graphs to Anumāna 2.

How can we formalise the application of step (3) to step (2), resulting in (4)? One of the things which is hard for any beginner to understand about Navya-Nyāya is exactly how this statement of pervasion and use of examples works – what distinguishes it from reasoning by analogy, in what sense it is *valid* reasoning? Authors such as Schayaer, Staal and others have characterised the third step in terms of predicate logic. For instance, Staal's occurrence relation,

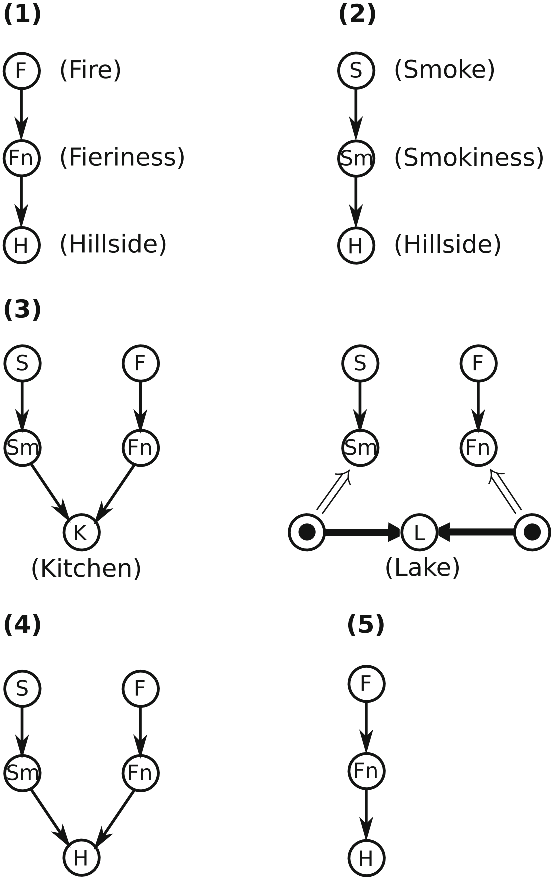


**Fig. 12.** Inference in Ganeri's graphs

$A$, links together the evidence, $h$, the site of inference, $p$ and conjecture, $s$: $A(h, p) \rightarrow A(s, p)$ (if $h$ is located in $p$ then $s$ must be located in $p$). [21] Whatever the notation, in order to know that the example presented is a reliable one, we must consider Diṅnāga's "statement with three characteristics". We need a diagrammatic means of knowing that the evidence appears in the site of inference (shown in step 2), in all things which are like the site of inference, and not in those things which are unlike that site. But step (3) only shows that the second and third characteristics are true of the particular examples. We also need to know that the site of inference and the corresponding positive example (*sapakśa*) are in the same likeness class with regard to to the target property, and conversely for the negative example (*vipakśa*). To be used in this context, Ganeri's graphs need to be extended since we have no way to describe these relationships. Das presents the same inference, fully specified, in pull-out Fig. 2 of his book [4]; we unfortunately don't have room to show this large diagram. The result is well-matched in that it corresponds, through Das' reading technique, with the Navya-Nyāya expression but, in our view, becomes too cluttered and abstruse to shed light on the key aspects of the inference (the three characteristics and the likeness and unlikeness classes). It seems that in order to construct a notation that exposes the distinctive features of Nyayān inference, a middle way between the expressiveness of Das and the elegance of Ganeri is needed.

## 5 Conclusion

We have collated the first survey of the use of diagrams in the Navya-Nyāya literature, and analysed the strengths and weaknesses of what has been done. Each author had their own goals in mind. Wada and Das produced a similar style of node-link diagram to illustrate their translations of Sanskrit texts, focusing on *vyapti* (the link between the evidence and that which we want to prove) and the structure of cognition. With the device of numbered contextual properties, Das' diagrams can be used to unambiguously reconstruct sentences in Navya-Nyāya. Ganeri produced his graphs to explore the Navya-Vaiśeṣika ontology, for which are an ideal match.

We believe that there is scope for further exploration of diagrammatic reasoning in logics which have intensional semantics and which focus on conceptual implication rather than propositional entailment, to use Mullick's terms. Our next step is to produce a notation which is iconic in the context of this kind of inference, which we will do by augmenting Ganeri's graphs with elements of Das' notation. The augmented notation needs to expose the most salient concepts involved in the application of the *parathanumāna* schema. A notation which does that must highlight the role of likeness and unlikeness classes through the links between the *sapakśa* (positive example), *vipakśa* (negative example) and the *pakśa* itself. It must make explicit the relation between the target and reason properties (i.e. which pervades which). In keeping with Ganeri's observations about the Vaiśeṣika ontology, the semantics of the notation should be graph-theoretic rather than model-theoretic. We will pursue this approach to explore the ways in which diagrams can illuminate both this distinctive style of reasoning and our own understanding of diagrammatic logic.

**Acknowledgements.** This work was done while Burton was at Jadavpur University as a Visiting Scholar. We thank Professor Prabal K. Sen and Professor Proyash Sarkar for some very helpful discussions on Navya-Nyāya and on the diagrams used by its practitioners.

## References

1. Ambrosio, C.: Iconic representations and representative practices. Int. Stud. Philos. Sci. **28**(3), 255–275 (2014)
2. Cheng, P.C.-H.: What constitutes an effective representation? In: Jamnik, M., Uesaka, Y., Elzer Schwartz, S. (eds.) Diagrams 2016. LNCS (LNAI), vol. 9781, pp. 17–31. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42333-3_2

3. Chi, R.S.Y.: Buddhist Formal Logic: A Study of Dignaga's Hetucakra and K'Uei-Chi's Great Commentary on the Nyayapravesa (Pt.1). Motilal Banarsidass Pub, Delhi (1990)
4. Das, R.C.: Navya-Nyayabhasapradipa of Mahesa Candra Nyayaratna. Utkal Studies in Philosophy. Utkal University Press, Bhubaneswar (2006)
5. Ganeri, J.: Indian Logic: A Reader. Routledge, London (2001)
6. Ganeri, J.: Indian logic and the colonization of reason. In: Ganeri, J. (ed.) Indian Logic: A Reader, pp. 1–25. Routledge, London (2001). Chap. 1
7. Ganeri, J.: Philosophy in Classical India: An Introduction and Analysis. Routledge, London (2001)
8. Goekoop, C.: The Logic of Invariable Concomitance in the Tattvacintāmaṇi. Springer, Netherlands (1967). https://doi.org/10.1007/978-94-010-3511-8
9. Guha, D.C.: Navya Nyāya System of Logic: Basic Theories and Techniques, 3rd edn. Motilal Banarsidass, Delhi (2016)
10. Gurr, C.: Effective diagrammatic communication: syntactic, semantic and pragmatic issues. J. Vis. Lang. Comput. **10**(4), 317–342 (1999)
11. Gurr, C., Tourlas, K.: Towards the principled design of software engineering diagrams. In: Proceedings of 22nd International Conference on Software Engineering, pp. 509–518. ACM Press (2000)
12. Hookway, C.: Truth, Rationality, and Pragmatism: Themes from Peirce. Clarendon Press, Oxford (2002)
13. Ingalls, D.: Logic in India. In: Ganeri, J. (ed.) Indian Logic: A Reader, pp. 110–117. Routledge, London (1955). Chap. 6
14. Matilal, B.K.: Perception. Clarendon Press, Oxford (1991)
15. Moktefi, A.: Is Euler's circle a symbol or an icon? Sign Syst. Stud. **43**(4), 597+ (2015)
16. Moktefi, A., Shin, S.-J.: A history of logic diagrams. In: Gabbay, D.M., Pelletier, F.J., Woods, J. (eds.) Handbook of the History of Logic, vol. 11, pp. 611–682. Elsevier, Amsterdam (2012)
17. Mullick, M.: Implication and entailment in Navya-Nyāya logic. J. Indian Philos. **4**(1–2), 127–134 (1976)
18. Schayer, S.: On the method of research into Nyāya. In: Ganeri, J. (ed.) Indian Logic: A Reader, pp. 102–109. Routledge, London (1933). Chap. 5
19. Sen, P.K., Chatterjee, A.: Navya-Nyāya logic. J. Indian Counc. Philos. Res. 77 (2010)
20. Short, T.L.: Peirce's Theory of Signs, 1st edn. Cambridge University Press, Cambridge (2009)
21. Staal, J.F.: The concept of paksa in Indian logic. J. Indian Philos. **2**(2), 156–166 (1973)
22. Stjernfelt, F.: On operational and optimal iconicity in Peirce's diagrammatology. Semiotica **2011**(186), 395–419 (2011)
23. Wada, T.: The analytical method of Navya-Nyāya. J. Indian Philos. **29**(5–6), 519–530 (2001)
24. Wada, T.: The Analytical Method of Navya-Nyāya. Gonda Indological Studies. Brill, Leiden (2007)

# Operations on Single Feature Indicator Systems

Atsushi Shimojima[1](✉) and Dave Barker-Plummer[2]

[1] Faculty of Culture and Information Science, Doshisha University,
1-3 Tatara-Miyakodani, Kyotanabe 610-0394, Japan
`ashimoji@mail.doshisha.ac.jp`
[2] CSLI/Stanford University, Cordura Hall,
210 Panama Street, Stanford, CA 94305, USA
`dbp@stanford.edu`

**Abstract.** A single feature indicator system (SFIS) is a signaling system where a representation carries information through a one-to-one correspondence of the "values" taken by its elements to those taken by a set of represented objects. The purpose of this paper is to demonstrate that many common diagrammatic systems are either SFISs or have SFISs as their semantic basis. We take as examples several familiar diagrammatic systems with seemingly diverse semantic systems (tables, charts, connectivity diagrams) and show the fundamental similarities among them that put them all under the concept of SFIS. We then explore different ways in which an SFIS is extended to a new, perhaps more expressive representation system. The paper paves the way to an account of the functional commonality and diversity of diagrammatic systems in terms of the operations that generate them from some basic systems.

## 1 Introduction

Investigations of the formal semantics of diagrammatic representations began with Shin's seminal work on Venn diagrams [17,18]. Since then further studies have been made of Euler diagrams, spider diagrams, Hyperproof diagrams, geometry diagrams, dot diagrams and coincidence grids among other representations, [1,3,7,8,11,12]. The formal semantics of so-called heterogeneous systems, combining diagrammatic and sentential representations, and multiple diagrammatic representations have also been investigated, [2,3,6].

The typical approach in this work is to define an abstract syntax for the diagrams, a model describing the subject matter for the diagram, and an interpretation function which determines how the diagram is to be read as making assertions about the model. The definitions of these three components are interdependent, and are specific to the particular diagrammatic system at hand.

Collectively, this work has made significant contributions to our understanding of logical properties of diagrammatic representation systems; however, the approach has been largely individualistic. The logical properties of each system

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 296–312, 2018.
https://doi.org/10.1007/978-3-319-91376-6_28

have been investigated on the basis of the semantic specification dedicated to that system. While the overall strategy for defining a formal semantics is shared, the specific mechanisms for implementing this strategy are idiosyncratic. It therefore is not adequate to account for systematic similarities between systems. For example, many diagrammatic representation systems support so-called free rides, and suffer from problems of over-specificity, [14]. The individualistic account does not have sufficient generality to describe when a system will have such properties, or what features of the system account for the presence of the property.

To address these issues, we previously proposed a *generic* approach to developing a theory of representation systems in general, with diagrammatic systems as a special case, [16]. Our approach involves giving an abstract characterization of a class of representation systems, and then investigating the properties of all members of the class in the abstract setting. By adopting this approach, we are able to short-circuit investigation of individual representation systems, and also to assign the responsibility for the possession of various properties of an individual representation system to its membership in the class. Specifically in [16], we formally characterized a particular class of representation systems that we call "Single Feature Indicator Systems (SFIS)" and showed that all SFISs share common properties, in particular that they admit a common collection of valid inference rules.

In the present paper, we extend our generic approach by exploring different ways in which an SFIS might be used as the basis for a new, perhaps more expressive representation system. We identify a set of operations that can be applied to SFISs to generate related, but functionally different representation systems. In doing so we begin to expand the universe of representation systems that can be described within our generic framework and show how different representation systems can be systematically related to one another.

## 2 The General Picture

In this and the next section we present an informal description of our view of representation systems and SFISs. The presentation here contains just enough information to serve as a foundation for the ideas presented here. For a more detailed view see [15,16].

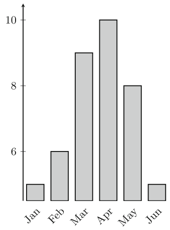
Our notion of representation system is designed to capture important semantic properties of a *representational practice* conducted by a group of people. A representational practice is a recurrent pattern in which people express information by creating a (typically proximal) object and extracting the information from it. In many cases, the information thus expressed is about a particular (distal) object or situation. We call a proximal object created in a representational practice a *representation*. A diagram is a special case of representation.

For example, a project leader may create the table in Fig. 1a to express information about the work schedules of four workers at a research project. Many people know how to extract the information expressed in this table and they do extract information from it. Here we see a representational practice conducted by the project leader and these people. We will refer to this representational practice as $\mathcal{R}_t$ and use it as the running example throughout this paper. The tables represent the schedules of the workers.

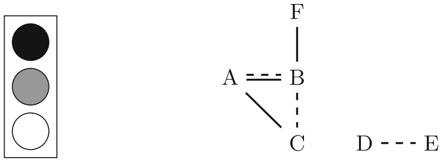
We can think similarly of the system of bar charts in Fig. 1b, representing the sales of a particular book over a period of months. In Fig. 1c, we see a system of traffic lights, representing a kind of traffic light whose individual lights may be illuminated, flashing or unilluminated, represented by white, gray and black circles respectively. Finally in Fig. 1d we see a connectivity diagram of a transit system.

| | Mon | Tues | Wed | Thurs | Fri |
|---|---|---|---|---|---|
| Atsushi | ✓ | ✓ | ✓ | ✓ | ✓ |
| Dave | ✗ | ✗ | ✗ | ✗ | ✗ |
| John | ✓ | ✗ | ✓ | ✗ | ✗ |
| Mike | ✓ | ✗ | ✓ | ✓ | ✗ |

(a) Scheduling Table.



(b) Bar Chart.



(c) Traffic Light. (d) Connectivity Diagram.

**Fig. 1.** Diagrams illustrating representation systems that either are SFISs themselves or modifications of SFISs.

Typically, a representational practice is governed by various constraints of different origins, and the effectiveness of the practice deeply depends on these constraints. They consist of *source constraints* concerning what kinds of symbols appear in a representation and how they are arranged, *semantic constraints* concerning what arrangement of symbols indicate what information, and *target constraints* holding on represented objects or situations.

The source constraints in $\mathcal{R}_t$ include the fact that each cell of the table contains one, and only one, of the symbols ✓ and ✗. This constraint is conventional, as it is based on essentially arbitrary syntactic stipulations among the participants of $\mathcal{R}_t$. Given the set of syntactic stipulations, however, further constraints on the arrangement of symbols naturally follow from them, due to spatial constraints on plane (e.g., the fact that there can be at most five checkmarks in a single row of the table). Our notion of source constraint covers these constraints too.

The target constraints include the facts that every worker either does, or does not, work on a particular day. They also include more idiosyncratic ones, such as the fact that at least one worker works on each day.

The semantic constraints include the fact that a cell has the symbol ✓ only if the relevant worker works on that particular day. As with syntactic stipulations, these semantic constraints are conventional in nature, and thus can be violated. Still, they are fairly strong regularities holding on $\mathcal{R}_t$, since the project manager, his workers, and other users know these conventions and try to comply with them to make their communication reliable and efficient.

The source constraints, semantic constraints, and target constraints governing a representational practice can be considered to make up a system, which we call a *representation system*.

## 3 Single Feature Indicator Systems

Having described our conception of representation systems, we now turn our attention to an important subclass of representation systems, Single Feature Indicator System (SFIS). They form the basis of many commonly occurring diagrammatic representation systems such as those presented in Fig. 1, including tables, bar charts and connectivity graphs. They are so fundamental to our way of codifying the world that new SFISs are constantly invented either out of blue or as modifications of existing SFISs.

### 3.1 The Structure of the Target Domain

The design of a SFIS starts from a set of *target issues*. This set reflects how we want to conceive of the given situation; it even determines which portion of the reality we take as our *target* in the first place. In the case of the system $\mathcal{R}_t$ of scheduling table (Fig. 1a) there are 5 × 4 target issues, corresponding to every combination of relevant worker and day, which may be written as:

Does Atsushi work on Monday?
Does Atsushi work on Tuesday?
. . .
Does Mike work on Friday?

A set of target issues determines the set of *subjects* that we are concerned with, as well as the range of values that these subjects may take. For the set of target issues listed above, the subjects are the 4 × 5 worker/day combinations, which may be written as "Atsushi on Monday", "Atsushi on Tuesday", . . . , "Mike on Friday"; the range of values is binary, consisting of "working" and "not working". This is the way the designer of the work scheduling tables wants to conceive the work schedule for a series of weeks.

We call a structure like this, in which there are a set of subjects and a range of values they uniquely take, a *feature*.

### 3.2 Designing the Source Representation

When the target object or situation is conceived on the basis of a feature, we have the opportunity to design a source representation which mirrors this structure.

Specifically, we can create another feature, this time in the source, whose sets of subjects and values stand in one-one correspondence to the sets of subjects and values of the target feature. For scheduling tables, the source feature has the set of subjects consisting of the $4 \times 5$ cells in a table, plus the set of values consisting of "containing ✓" and "containing ✗". The cells in the source feature stand in a one-one correspondence to the worker/day combinations in the target feature, and the binary values "containing ✓" and "containing ✗" similarly correspond to the binary values "working" and "not working."

We call the feature consisting of the sets of source subjects and source values *the source feature*. Note that the values taken by the various subjects in the source feature are *independent*. That is, as far as the syntactic stipulations and spatial constraints are concerned, a source subject in a given diagram can take any value without consideration of the values taken by other source subjects. We call this the *independence condition*.

### 3.3 Semantic Conventions

The semantic conventions can be specified on the basis of the twofold correspondences holding between the source feature and the target feature. In the case of $\mathcal{R}_t$, they consist of the constraints of the following form, where $X$ ranges over the names of the workers and $Y$ over the names of the five days of the week:

(1) If the cell in the row labeled by the name $X$ and the column labeled by the name $Y$ has a ✓, then the person bearing $X$ works on the day bearing $Y$.
(2) If the cell in the row labeled by the name $X$ and the column labeled by the name $Y$ has a ✗, then the person bearing $X$ does not work on the day bearing $Y$.

Generally, establishing the basic semantic conventions for a SFIS involves three steps: (1) specifying which subject in the source feature corresponds to which in the target feature, (2) specifying what value in the source feature corresponds to which value in the target feature, and (3) making it a rule that a subject in the source feature takes a value only if the corresponding subject in the target feature takes the corresponding value.

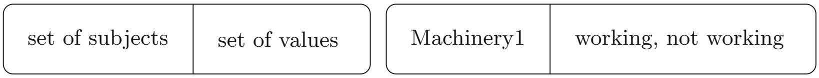
Informally, a representation system is a *Single Feature Indicator System* (SFIS) if it satisfies the following conditions:

1. Due to the system's target constraints, every target subject is constrained to take a unique target value (existence of a target feature),
2. Due to the system's source constraints, every source subject is constrained to take a unique source value (existence of a source feature),
3. The source feature satisfies the independence condition, and
4. The source and target features are linked by the form of semantic conventions specified above.

The simplest possible SFIS has a single subject in the target feature taking on a binary value and a corresponding single subject in the source feature also with a binary value. An example of such a SFIS might have a particular piece of machinery as its sole target subject, taking the value "working" or "not working". Its sole source subject is then a particular lamp on a control panel, taking the value "illuminated" or "not illuminated". We call such a simple signaling system an *atomic* single feature indicator system. Atomic SFIS are quite common, and we will take them as our starting point in the consideration of the operations that can be performed on SFISs.

## 4 Notation

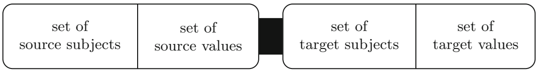
In what follows, we will use diagrams to describe SFISs. An individual feature within an SFIS, will be represented schematically as a divided box, with a set of subjects to the left, and a set of values to the right, as in Fig. 2a. A specific example, with a feature of a machine that is either working or not, is shown in Fig. 2b.



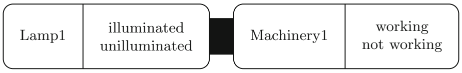
(a) A Generic Feature (b) An Example Feature

**Fig. 2.** Notation for features

An SFIS then, can be represented as two features, source to the left and target to the right, with a semantic link between them. Since we are not focussed on semantic links in this paper, we will depict (and treat) it as a black box between the two features. Figure 3a shows the general form of a SFIS. A concrete SFIS involving a single lamp depicting the state of some machine is shown in Fig. 3b.



(a) Generic SFIS



(b) An Example SFIS

**Fig. 3.** Notation for SFISs

## 5 Operations: Subject Union

Given the lamp-machinery example just presented, it is very natural to imagine a control panel representing the status of the machinery in a complex plant, a factory for example. In this situation, there are many pieces of machinery, each either operating or not. The status of each piece of machinery can be represented by a lamp on a complex control panel. If there are three distinct subsystems in the factory, then these can be represented using three lamps, any of which is independently illuminated or unilluminated.

We can analyze such a system as formed from three underlying SFISs by forming the union of the subjects, as depicted in Fig. 4. In this figure, we can see that the three atomic SFISs depicted on the left have been combined into a single SFIS on the right. Each feature in the new system has the union of subjects from the component SFISs. The operation of subject union is applicable to SFISs only when they share the same sets of source and target values.

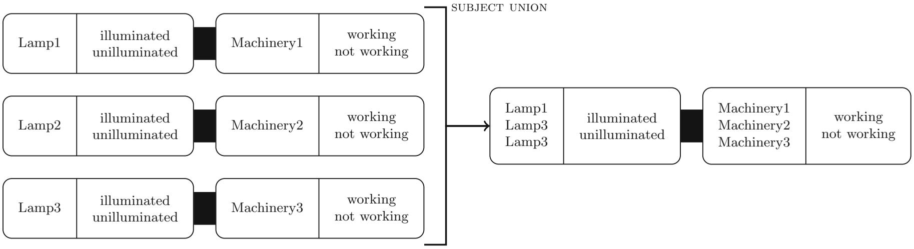


**Fig. 4.** An example of the subject union operation

A more interesting use of subject union is commonly found in connectivity maps. For example, we may want to design a transit map for the subway system for a city. In such a system, we can think of the target subjects as pairs of stations, and the associated values as "directly connected" (*dc*) or "not directly connected" ($\neg dc$). The simplest such feature is concerned with two stations `A` and `B`, say, and hence one subject—the pair (`A`,`B`).[1] One common way to express such connection, is with stations represented by nodes on the plane, and direct connectivity represented by the presence (*edge*), or absence ($\neg edge$), of an edge between two nodes. This forms an atomic SFIS.

Figure 5 shows this situation. On the left are three atomic SFISs representing direct connectivity between different pairs of stations. The SFIS depicted at the top describes the connectivity relation between stations `A` and `B`. In the sample diagram to its left, these stations are depicted as directly connected by the presence of an edge between the nodes "A" and "B". The other two atomic SFISs have similar features involving different pairs of nodes and stations. We

---

[1] We assume that stations are not connected to themselves, and that connections are bidirectional.

show sample diagrams where "A" is directly connected to "C" and that "B" is not directly connected to "C".

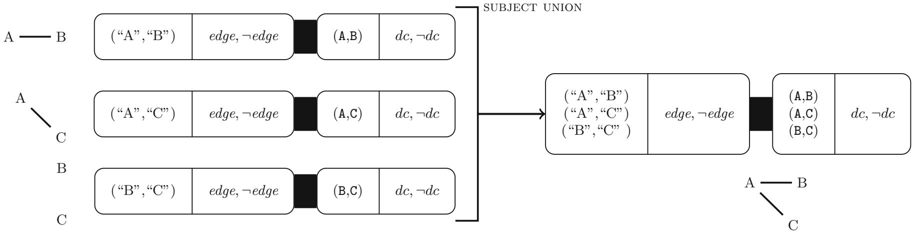


**Fig. 5.** An example of the subject union operation for connectivity

The representation system to the right has all three pairs of stations as target subjects, and all three pairs of nodes as source subjects, with the set of source values inheriting the same semantics.

## 6 Operations: Value Multiplication

Imagine now that our transit system contains multiple different subway lines, and that we therefore have a collection of SFISs of the type that we just described, each allowing the representation of the connectivity using one particular line. The SFIS on the right of Fig. 5 is the system for representing the connectivity between stations on line $P$, say.

We would like to overlay these SFISs into a single representation system to allow riders to determine how to travel between any two stations in the network possibly by changing trains. We call this overlaying operation *value multiplication*.

An example of this is depicted in Fig. 6. In this situation, we have two SFISs sharing the same set of subjects. The values for each SFIS are analogous, but not the same. In one source feature we draw solid edges in the representation to signal that the stations are directly connected on line $P$, and in the other we draw dashed edges to signal that the stations are connected on line $Q$. These SFISs on the left of the figure represent possible journeys carried out exclusively on one line. Examples of concrete representations are shown to the far left of the figure.

The system depicted on the right is formed by taking the cartesian product of the value sets. So a value in this new system is a pair, consisting of one possible value drawn from each of the component sets. It is easy to show that this new representation system is itself a SFIS. The combined concrete representation shows the overlay of the two component representations.

Systems based on value multiplication are quite common. Imagine for example, an extension of bar charts where the months whose values exceeded last

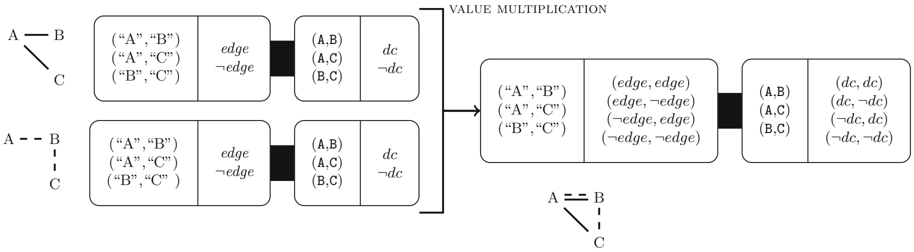


**Fig. 6.** Value multiplication

years sales are colored green, and those that did not are colored red. Here a multiplied system is produced from the bar chart system together with another system sharing the same subjects but having different sets of source values ("red" and "green") and target values ("exceed" and "¬exceed"). The source values of the multiplied system would consist of pairs of bar heights and colors, which correspond to pairs of sales numbers and "exceed/¬exceed" values.

## 7 Operations: Neutralization

With a SFIS we are required to convey complete information about the target situation that specifies the value of each individual target subject. However, we are not always in a position in which complete information is available.

Consider a situation where you are observing the author of the scheduling table in Fig. 1a as it is being constructed. Perhaps all of the row and column labels are present, but the author has not yet filled in all of the cells. Such a diagram is not a representation in the scheduling table system $\mathcal{R}_t$, since not every source subject takes on a value from ✓, ✗. However, such a representation does carry *partial* information about the target that it describes. We can see, perhaps, that John will be working on Monday, but whether or not Atsushi is also working that day is not expressed in the diagram.

We can model such a situation by noticing that the blank space in the source representation serves as an additional value that source subjects can take on, but that this value is associated with the disjunction of all the other values in the target feature. In the case of the scheduling table system $\mathcal{R}_t$, we will denote this disjunctive value by $\bigvee\{working, not\ working\}$. We call such a value a "neutral" value, and a resulting system a *Single Feature Indicator System with neutrality*, or a *neutralized* SFIS. We give a formal account of neutralized SFIS in [16].[2] Unlike the cases of subject union and value multiplication, the result of neutralizing an SFIS is a new representation system which is not itself an SFIS.

---

[2] In [16] we give a different account of neutral values as not referring to any value in the target system. These two accounts are equivalent, but the account here allows us to generalize the idea of complete neutrality.

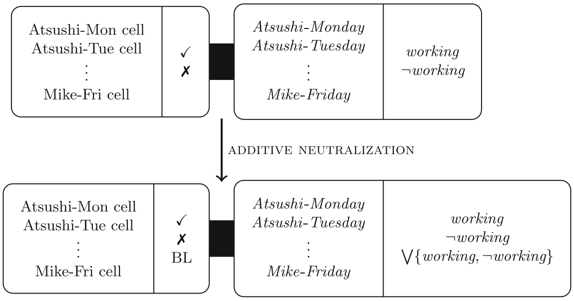


**Fig. 7.** Example of additive neutralization. The source value, BL, has no semantic significance as it corresponds to the neutral target value, $\bigvee\{working, \neg working\}$.

Naturally, neutralized SFISs do not only occur on the way to completing the construction of an SFIS, but they are useful when information about the domain is, perhaps necessarily, partial. Consider the traffic light system, where individual lights might be either "illuminated", "unilluminated" or "flashing". When we don't know the state of a light in the target situation, we might use a new color in the source to remain neutral about the issue. The resulting diagram expresses partial information about the world.

More generally, neutralization might not necessarily indicate a complete absence of information. Rather, newly added source values, might be introduced to indicate partial, but not totally absent information. In the traffic light system above, we might introduce a way to represent the fact that a light is either illuminated or unilluminated (i.e. not flashing) by introducing a source value associated with the target value $\bigvee\{illuminated, unilluminated\}$.

The case above is an example of what we will call *additive* neutralization (Fig. 7). A source value is added in the process of neutralization, with the stipulation that it corresponds to a disjunctive target value.

Sometimes an existing representation system can be neutralized by making an existing source value correspond to a neutral target value. We call the process of modifying a representation system in this way *subtractive* neutralization. The reader familiar with Euler diagrams may notice that the currently standard system of Euler diagrams (e.g., [6]) is the result of subtractive neutralization applied to the "older" Euler system with so-called "existential import" (e.g., [9]). In this case, the minimal regions (or zones) of Euler diagrams are source subjects and the sets represented by them are target subjects. In both systems, if a minimal region is not present in a diagram, it means that the represented set is empty. If a minimal region is present, however, it does *not* mean the non-emptiness of the represented set in the newer system, while it does so in the older system. In terms of SFISs the set of source values $\{present, \neg present\}$ in the older system correspond to the set of target values $\{\neg empty, empty\}$, whereas that in the newer system corresponds to the set of target values $\{\bigvee\{empty, \neg empty\}, \neg empty\}$, making source value "*present*" correspond to $\bigvee\{empty, \neg empty\}$. Thus, we can see the newer system as the neutralized version of the older system, although neutralization in this case consists in making an existing source value correspond to a neutral target value, rather than adding a new source value.

## 8 Operations: Conditionalization

The definition of SFIS requires that different source subjects take their values independently of one another because of the independence condition on the source feature. In the example of the work scheduling system $\mathcal{R}_t$, the fact that the Atsushi-Mon cell has a checkmark does not affect, as far as syntactic and spatial constraints are concerned, whether the Atsushi-Fri cell has a checkmark, whether the Dave-Mon cell has a checkmark, or whether the Dave-Fri cell has a checkmark, for example. However, when stipulations constraining the values assumed by some source subjects on the basis of other subjects' values are added to a SFIS, we say that the new system results by the operation of *conditionalization.*

For example, when examining a working traffic light, it is impossible (hopefully) for both the green and red lights to be on simultaneously. In this case there is a constraint in the target domain. The fact that one subject, the red light, assumes the value "*illuminated*", precludes the fact that the green light can also take this value. This constraint is a *local* constraint. It says that the value of some target subject constrains the value of some other target subject, but leaves the remaining subjects unconstrained.

When representing proportional data we often encounter global constraints. Consider for example the results of an single vote election with four candidates. Once the first candidate is known to have received 25% of the votes, then it is impossible for any of the other candidates to receive 80% of the votes. This inter-candidate constraint is a *global* constraint on the target system, namely that the total of votes cast must sum to 100%. Fixing the value for any candidate, constrains *all* of the values for the other candidates.

These examples are different in interesting ways. If the target feature for the traffic light system is constrained in this way, but we continue to use pencil and paper diagrams for representing the traffic lights, then the structural constraint in the target feature is not projected into the representation. We can still draw diagrams in which the red and green lights are both represented as on, even though we know that this does not represent a permissible target situation. In order to work with such a representation, then users of the system must keep track of the fact that some diagrams do not represent possible situations, and reason with them accordingly. If we ameliorate this problem by implementing a simple computer-based editor for these diagrams which tracks the target constraint, it is a case of conditionalization. In this system the revised source feature prevents the construction of representations of impermissible situations.

By contrast, we can choose to represent proportional data using a representation such as a pie chart. Even when drawn with a pencil on paper, if the angles of the slices of the pie accurately reflect the proportions of the represented data, then it becomes impossible to draw a chart that represents an impossible situation. This ability to prevent the representation of impossible data even when using very flexible tools such as pencil and paper, accounts for much of the wide adoption of pie charts as a system for representing proportional data. The system is a naturally conditionalized system, so to speak.

## 9 Operations: Uplifting

When we combine SFISs through subject union or value multiplication, we typically adopt natural (but essentially arbitrary) syntactic stipulations on the way source subjects appear in diagrams, and that creates the possibility of perceptually available *derived meaning.*

For example, the cells of a scheduling table in $\mathcal{R}_t$ are organized into a grid so that the rows share a person label and the columns share a day label. This results in the possibility of observing facts such as "Mike's row has three checkmarks", which indicates the fact that Mike works three days this week. The syntactic constraints for this organization of cells makes a new representational practice available, where one counts the number of checkmarks in a particular person's row and reads off the number of days that person works in the week. The fact that the number of checkmarks in a row indicates the number of days that a person works is a consequence of the semantics of the underlying system. This is an instance of what is called *derivative meaning relation* [14]. We call the move from a system involving basic meaning relations to one using derivative meaning relations, *uplifting.*

Figure 8 shows the systems involved in our example. The leftmost, cell-based SFIS serves as the basis for the uplifting of a row-based representation system (to the right).

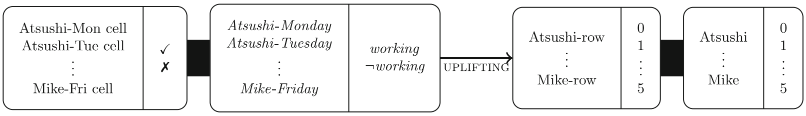


**Fig. 8.** An example of uplifting and system combination. A SFIS is uplifted from another SFIS and combined with it to make a semantically richer system.

In this particular example, the uplifting operation produces a new SFIS where the source subjects are the rows of cells, and the values are integers from 0 to 5. In general, there may be many ways to read derived information from a representation system, and not all uplifting operations based on this derived information will result in an SFIS. For example, we can uplift the cell-based representation system to a new representation where the subjects are pairs of columns of the table, and the value associated with a subject is the difference in the number of checks in the columns (indicating the difference in the number of people working on the corresponding days.). See Fig. 9. The resulting system is not an SFIS since the independence condition fails. Once the difference in the number of workers between Monday and Tuesday is known, and the corresponding value for Tuesday and Wednesday is also known, then the difference between Monday and Wednesday is determined.

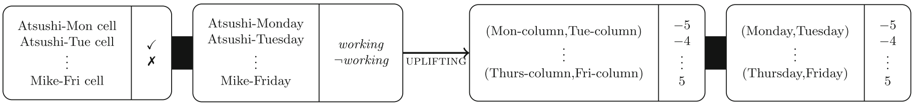


**Fig. 9.** An alternative uplifting of the cell-based scheduling table system

Uplifting can be also applied to the SFISs of bar charts and connectivity maps in Fig. 1. Due to syntactic stipulations requiring bars to be aligned at bottom, one may compare the vertical positions of the tops of two bars to compare the sales in the relevant months. This practice is based on an uplifted SFIS, where pairs of bars correspond to pairs of months and the differences of the vertical positions of bar tops correspond to the differences of the sales of months. Bars are also required to be even in their width and horizontal positions, so one may also uplift a SFIS that lets the shape formed by the entire set of bars indicate whether the sales trend is rising, falling, flat, or peaked.

The multiplied SFIS of connectivity maps in Fig. 1d can be uplifted to a new SFIS addressing the general connectivity of two stations in the subway system, rather than the direct connection via an individual subway line. For example, we see a *path* between the nodes “A” and “F” in Fig. 1 but not between nodes “A” and “D”. These facts indicate, respectively, that stations `A` and `F` are connected in the subway system and that `A` and `D` are not connected. If our basic SFIS had more pairs of stations as target subjects, it could be uplifted to a system concerning the number of stations on the shortest path between two stations, with non-negative integer values for connected pairs of stations and “Not Applicable” for non-connected pairs of stations.

The uplifting of a SFIS is possible because of the general fact that new meaning relations can be logically derived from more basic ones. We refer the reader to [14] for illustration of how prevalent meaning derivation is among diagrammatic systems. The uplifting operation offers a logical account of what has been known as “global reading” [13] or “macro reading” of diagrams [19]. Through uplifting, the granularity of source and target subjects is typically enlarged (e.g., from individual cells to entire rows, or from individual bars to a group of bars). Thus, reading based on the uplifted system typically requires attention to larger structures in diagrams than that based on the base SFIS. Hence the contrast between global and local reading. On this account, learning how to read diagrams often consists in the acquisition of not only a base SFIS but various uplifts from it. Expertise of reading a particular class of diagrams (e.g. meteorological maps

[10] or stock-price charts [4]) then depends on the varieties of uplifts that one has acquired from the relevant base system.

## 10 Operations: System Combination and Elimination

Individuals in a representational practice might behave differently with regard to uplifted systems in different situations.

In some situations individuals in the practice work with a complex combination of the different systems simultaneously. Experts of reading a class of diagrams (Sect. 9) probably have acquired such complex systems. Also in the case of familiar scheduling tables, we can imagine a scheduler noticing that only one person is working on Tuesday (column-based system), that Dave is not working on Tuesday (cell-based system) and that since Dave is not working at all this week (row-based system), perhaps he should be scheduled to work on Tuesday. We describe this more complex system using the operation of *system combination* in which a complex representation system is formed from the simultaneous use of multiple systems sharing the same underlying SFIS.

We can view this as an example of *heterogeneous reasoning* [2,3,6] as it combines multiple representation systems with different semantic makeups. The systems involved however do not look obviously heterogeneous since they consist of a particular SFIS plus systems uplifted from it, and hence work on identical representation tokens (the same scheduling table in the present example). Heterogeneous reasoning in this extended sense is closely related to the idea of *aspect-shifting*, which has been discussed as a pervasive and powerful feature of diagrammatic representation systems [5,8,14].

In other situations, some aspects of a combined system may be sufficient for the kinds of issues that the individuals have to address in their practice. A typical case is where the original primitive system becomes irrelevant to the practice over time and only the uplifted systems come to be used. Take the case of bar charts. When presented a bar chart like the one in Fig. 1b, we are often more interested in the general trend of book sales or the relation of the sales of different months, than the exact number of sales indicated by individual bars. At certain times, that becomes our sole interest, and it becomes an established practice, for both chart producers and viewers, to use uplifted SFISs only. In this case we can conceive of the operation of *system elimination* as a mechanism by which a complete sub-system of representations is removed from the representational practice.

What, then, is it that we have multiple systems *combined*, as opposed to having them separately? How does the combined system let people work with single selected sub-systems at some times while being able to integrate pieces of information from multiple sub-systems at other times? Answering these question clearly requires more structures to be added to the model of representation systems that we have posited for this paper. The task has to be left for future work.

## 11 Conclusion

The operations that we have explored in this paper show that many diagrammatic systems, despite their apparent functional diversity, are based on rather simple signaling systems: they are either instances of SFISs themselves or their syntactic and semantic derivatives. The operations of subject union and value multiplication show that apparently complex systems are constructions from atomic SFISs, making up new SFISs themselves. The operations of additive and subtractive neutralization identify an important class of SFIS-derivatives, which allow the expression of partial information in diagrams where some of the values of target subjects are left undetermined. The operation of conditionalization identifies another class of SFIS-derivatives that track constraints on targets by spatial or physical constraints on the way source subjects take values. The operation of uplifting shows how a single SFIS can serve as the semantic basis of multiple derivative systems that capture target subjects in different granularities.

The final two operations, i.e. system combination and elimination, require more detailed characterization in future work. Yet, the operation of combination does suggest an account of the capacity of diagrammatic systems to support reasoning that looks at targets from different perspectives; the operation of elimination provides an account for the change of syntax of diagrammatic systems as the result of cultural selection of particular perspectives in combined systems.

In [16], we have proposed the generic approach to the logical study of diagrammatic representations as a complement to the conventional individualistic approach. While the approach in [16] focuses on demarcating the class of SFISs and emphasizes the similarities of diagrammatic systems falling in this class, the present paper emphasizes *variations* of diagrammatic systems by characterizing various operations applied to SFISs, which generate different types of SFISs and non-SFISs.

The importance of this extended generic approach consists in the fact that these operations are systematically correlated with functional modulations of the systems involved. We have already suggested several instances. One was concerned with the neutralization operation. By nature, a SFIS suffers from severe over-specificity problem, which prevents a SFIS-diagram from expressing information without specifying the value of every individual target subject. The neutralization operation seems to alleviate this problem significantly. Another instance was the uplifting operation. It opens up the possibility of expressing macro-level pieces of information, while the basic SFIS can express only piecemeal information about individual target subjects. While neutralization alleviates the over-specificity problem of an original SFIS, it seems to significantly reduce this potential of uplifting, since it weakens the semantic link between the source and target features in the original system.

Generally, diagrammatic systems lose, preserve, and acquire functions in a systematic manner when different operations are applied to them. So, by investigating varieties of operations applicable to SFISs and other systems, and by specifying the functional modulations they give rise to, it is possible to account for the functional diversity of diagrammatic systems in a progressive manner. We believe that the present paper opens up one avenue in such an endeavor, although its coverage is admittedly limited given the whole range of system-operations that our culture has invented.

## References

1. Barker-Plummer, D., Swoboda, N.: Reasoning with coincidence grids—a sequent-based logic and an analysis of complexity. J. Vis. Lang. Comput. **22**(1), 56–65 (2011)
2. Barker-Plummer, D., Swoboda, N., Murray, M.D.: An example HyperVenn proof. In: Dwyer, T., Purchase, H., Delaney, A. (eds.) Diagrams 2014. LNCS (LNAI), vol. 8578, pp. 51–53. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44043-8_8
3. Barwise, J., Etchemendy, J.: Heterogeneous logic. In: Glasgow, J., Hari Narayanan, N., Chanrasekaran, B. (eds.) Diagrammatic Reasoning: Cognitive and Computational Perspectives, pp. 211–234. MIT Press/AAAI Press, Cambridge/Menlo Park (1995)
4. Bulkowski, T.N.: Encyclopedia of Chart Patterns, 2nd edn. Wiley, Hoboken (2005)
5. Giaquinto, M.: Visual Thinking in Mathematics: An Epistemological Study. Oxford University Press, Oxford (2007)
6. Hammer, E.: Logic and Visual Information. CSLI Publications, Stanford (1995)
7. Howse, J., Molina, F., Taylor, J., Kent, S., Gil, J.Y.: Spider diagrams: A diagrammatic reasoning system. J. Vis. Lang. Comput. **12**, 299–324 (2001)
8. Jamnik, M.: Mathematical Reasoning with Diagrams. CSLI Publications, Stanford (2001)
9. Lemon, O., Pratt, I.: Spatial logic and the complexity of diagrammatic reasoning. Mach. Graph. Vis. **6**(1), 89–108 (1997)
10. Lowe, R.K.: Diagram prediction and higher order structures in mental representation. Res. Sci. Educ. **24**, 208–216 (1994)
11. Miller, N.: Euclid and His Twentieth Century Rivals: Diagrams in the Logic of Euclidean Geometry. CSLI Publications, Stanford (2007)
12. Mineshima, K., Okada, M., Takemura, R.: A diagrammatic inference system with Euler circles. J. Log. Lang. Inf. **21**(3), 365–391 (2012)
13. Ratwani, R.M., Trafton, J.G., Boehm-Davis, D.A.: Thinking graphically: connecting vision and cognition during graph comprehension. J. Exp. Psychol.: Appl. **14**(1), 36–49 (2008)
14. Shimojima, A.: Semantic Properties of Diagrams and Their Cognitive Potentials. CSLI Publications, Stanford (2015)
15. Shimojima, A., Barker-Plummer, D.: The Barwise-Seligman model of representation systems: a philosophical explication. In: Dwyer, T., Purchase, H., Delaney, A. (eds.) Diagrams 2014. LNCS (LNAI), vol. 8578, pp. 231–245. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44043-8_25
16. Shimojima, A., Barker-Plummer, D.: A generic approach to diagrammatic representation: the case of single feature indicator systems. In: Jamnik, M., Uesaka, Y., Elzer Schwartz, S. (eds.) Diagrams 2016. LNCS (LNAI), vol. 9781, pp. 83–97. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42333-3_7

17. Shin, S.-J.: A situation-theoretic account of valid reasoning with Venn diagram. In: Barwise, J., Gawron, J.M., Tutiya, S. (eds.) Situation Theory and Its Applications, pp. 581–606. CSLI Publications, Stanford (1991)
18. Shin, S.-J.: The Logical Status of Diagrams. Cambridge University Press, Cambridge (1994)
19. Tufte, E.R.: Envisioning Information. Graphics Press, Cheshire (1990)

# The Observational Advantages of Euler Diagrams with Existential Import

Gem Stapleton[1](✉), Atsushi Shimojima[2], and Mateja Jamnik[3]

[1] Centre for Secure, Intelligent and Usable Systems,
University of Brighton, Brighton, UK
`g.e.stapleton@brighton.ac.uk`
[2] Doshisha University, Kyoto, Japan
`ashimoji@mail.doshisha.ac.jp`
[3] Department of Computer Science and Technology,
University of Cambridge, Cambridge, UK
`mateja.jamnik@cl.cam.ac.uk`

**Abstract.** The ability of diagrams to convey information effectively in part comes from their ability to make facts explicit that would otherwise need to be inferred. This type of advantage has often been referred to as a *free ride* and was deemed to occur only when a diagram was obtained by translating a symbolic representation of information. Recent work generalised free rides to the idea of an *observational advantage*, where the existence of such a translation is not required. Roughly speaking, it has been shown that Euler diagrams *without* existential import are *observationally complete* as compared to symbolic set theory. In this paper, we explore to what extent Euler diagrams *with* existential import are observationally complete with respect to set-theoretic sentences. We show that existential import significantly limits the cases when observational completeness arises, due to the potential for overspecificity.

## 1 Introduction

Diagrams are often seen as a useful tool in aiding our understanding of information, particularly in contrast to symbolic or textual notations. One of many reasons for this can be attributed to the ability of diagrams to convey facts in accessible manner, including facts that would otherwise need to be inferred from alternative representations. Such facts can be thought of as *observable* from the diagrammatic representation but *inferrable* from the alternative representation.

Previously, we introduced the theory of *observational advantages* [16], generalising the idea of a free ride [14]. In the case of free rides, one starts with a collection of statements, say $Q \subseteq P$ and $P \cap R = \emptyset$, and translates them into a semantically equivalent diagrammatic form, such as in Fig. 1. The translation must ensure that the original statements can all be *observed* from the diagram. It can readily be seen, or observed, from Fig. 1 that $Q \subseteq P$ (through curve containment) and $P \cap R = \emptyset$ (through curve disjointness). In addition, it can also be observed that $Q \cap R = \emptyset$, again through curve disjointness. This information,

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 313–329, 2018.
https://doi.org/10.1007/978-3-319-91376-6_29

$Q \cap R = \emptyset$, is a free ride from the diagram given the original statements. In the case of observational advantages, the requirement for a translation is removed: the two representations must only be semantically equivalent, implying that all free rides are observational advantages but not vice versa.

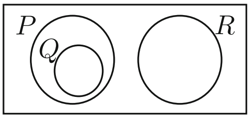


**Fig. 1.** No existential import and a free ride.

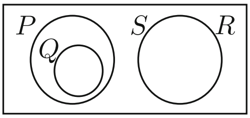


**Fig. 2.** Overspecificity issues.

The observational power of diagrams has long been recognised, with Hyperproof incorporating an observation-style inference rule in a proof system involving both diagrams and first-order logic [2]. Dretske's work, commonly described as "somebody's seeing that something is the case" [5], informed the development of the Euler/Venn inference system, where the authors called for the distinctive treatment of observation [18]. By formalising this insight into the benefit of diagrams, we can identify the set of statements that are *observable*, as opposed to *inferrable*, from a given statement.

In [16] we presented a formal framework for studying observational advantages and applied it to Euler diagrams (without existential import [3][1]) and set-theoretic sentences, limited to making subset and equality assertions. We proved that Euler diagrams were *observationally complete*: given any finite collection of set-theoretic sentences, $S$, from the class just described, there exists an Euler diagram, $d$, from which any set-theoretic sentence, $\sigma$, inferrable from $S$ can be observed. This is a significant result, because Euler diagrams are widely studied from the perspective of inference [4,10,13], demonstrating that diagrams can be rich in observational advantages. An obvious question arises: if Euler diagrams are instead taken to have existential import, are they still rich in observational advantages? Given that many approaches exist to asserting non-emptiness (see [11]), including existential import, it is important to answer this question.

We extend our previous work to Euler diagrams *with existential import* and set-theoretic sentences which can also express the *non-emptiness of a set* and *non-subset relations between sets*. This allows us to show that there are restricted instances of when a finite set, $S$, of set-theoretic sentences has an equivalent Euler diagram, $d$, with existential import from which any set-theoretic sentence, $\sigma$, inferrable from $S$ can be observed. This is due to the *overspecificity* of Euler diagrams *with* existential import.

---

[1] In Euler diagrams without existential import, zones can represent empty sets. By contrast, under existential import all zones in the diagram represent non-empty sets [7]. Peirce denotes non-emptiness of a set with ⊗-sequences [12] (also used by Shin [15] and further developed by Choudhury and Chakraborty [4]). Other notations use graphs to denote elements in sets [6,8,9].

As an example, consider $Q \subseteq P$, $P \cap R = \emptyset$, $P \backslash Q \neq \emptyset$, $R = S$ and $Q \not\subseteq S$. One possible visualisation of this using Euler diagrams with existential import is in Fig. 2. Clearly the four statements are represented in the diagram; focusing on $P \backslash Q \neq \emptyset$, this corresponds to the zone inside just the curve $P$ (so outside $Q$, $R$ and $S$), which represents a non-empty set due to the existential import assumption now placed on the diagram's semantics. However, the diagram represents *too much information* such as $S \neq \emptyset$: forcing zones to represent non-empty sets leads to this overspecificity problem [14,17]. Indeed, under the existential import semantics, there is *no single Euler diagram* that represents just these four statements and nothing more. In the rest of this paper, all Euler diagrams are assumed to be interpreted under the existential import semantics.

We provide the necessary and sufficient conditions under which a set of set-theoretic sentences, formed using $\subseteq$, $\not\subseteq$, $=$ and $\neq$, can be visualised as an Euler diagram *with* existential import. We show that such a diagram is observationally complete. The conditions demonstrate that existential import may not only restrict the existence of an observationally complete diagram, but may prevent a semantically equivalent diagram to exist at all. Our results show that Euler diagrams with existential import suffer from overspecificity, which hugely limits their advantages over competing notations. This insight sets this paper apart from earlier work: it is the first to formally reveal that diagrams can have (substantial) limitations in exchange for the power to express a wider variety of information (such as $\not\subseteq$ and $\neq$). Consequently, designers and users of diagrams should pay careful attention when defining their syntax and semantics if one of their goals is to harness their observational power.

The paper is structured as follows. Section 2 discusses the idea of a meaning-carrying relationship and its role in observation. The syntax and semantics of Euler diagrams and the fragment of set theory that we consider are given in Sect. 3. We provide results on the model theory of these two systems in Sects. 4 and 5 respectively, which are the necessary basis for understanding the limitations of Euler diagrams with existential import. Section 6 establishes the limited set of cases when observational completeness arises. We discuss these results and their implications in Sect. 7 and conclude in Sect. 8.

## 2 Observation and Meaning-Carriers

Central to the notion of observability is an understanding of how a representation of information conveys meaning through *meaning-carrying relationships*, which is discussed at some length in [16]. Here, due to space constraints, we provide a brief discussion along with various definitions from [16] that are essential for the remainder of this paper. As the context of this paper is on set theory and Euler diagrams, we provide examples from those domains.

A meaning-carrying relationship is a relation on the syntax of a statement that carries semantic value, evaluating to either *true* or *false*. This is similar to Shin's notion of a *representing fact* in her seminal work on Venn diagrams [15].

Meaning-carriers play an important role in both her work and ours. In set-theoretic sentences, such as $P \subseteq Q$, there are single meaning-carrying relationships. In $P \subseteq Q$, the meaning-carrier is that $P$ is written to the left of $\subseteq$ and $Q$ to the right. Likewise, any set-theoretic statement formed using $\not\subseteq$, $=$, and $\neq$ has just a single meaning-carrier: the set written on the left has the asserted relationship with that written on the right.

A single Euler diagram can have numerous meaning-carrying relationships which are given by the spatial relationships between the curves. In Fig. 2, the curves $S$ and $R$ are on top of one another, asserting that $S = R$. These meaning-carriers in the diagram give rise to the observable set-theoretic sentences. From Fig. 2, we can observe $Q \subseteq P$, $Q \cap R = \emptyset$, and $S = R$, amongst many other things. Due to existential import, we can also observe $S \neq \emptyset$ and $P \backslash Q \neq \emptyset$.

One has to understand which syntactic relationships are meaning-carriers in order to define observability. In particular, one statement, $\sigma_1$, is observable from another, $\sigma_2$, if some meaning-carrying relationship in $\sigma_2$ corresponds directly to $\sigma_1$. For example, the containment of one curve, $Q$, by another, $P$, in an Euler diagram (Fig. 2) is a meaning-carrier, allowing us to observe a subset relationship: $Q \subseteq P$. Likewise, when we have existential import, the presence of a region is a meaning-carrier: it represents a non-empty set. So, a set arising from a region, such as the zone inside $P$ and $Q$ in Fig. 2, is non-empty: $P \cap Q \neq \emptyset$. We can also observe from the diagram $P \cap Q = Q$, since the region which represents $P \cap Q$ happens to be exactly the same region as that which represents $Q$.

Importantly, observability must respect semantics too: if a statement is observable then it must be semantically entailed (i.e., the observed statement must be true whenever the statement from which it is observed is true). We will define when a set-theoretic sentence can be observed from an Euler diagram later[2]. For now, we assume this definition is given and present a general definition of observability from a *set* of statements:

**Definition 1.** *Let $\Sigma$ be a finite set of statements and $\sigma_o$ be a statement. Then $\sigma_o$ is **observable** from $\Sigma$ iff $\sigma_o$ is observable from some statement, $\sigma$, in $\Sigma$. The set of statements that are observable from $\Sigma$ is denoted $\mathcal{O}(\Sigma)$* [16].

We can now define what it means to be observationally complete:

**Definition 2.** *Let $\Sigma$ and $\Sigma_{\models}$ be finite sets of statements. Then $\Sigma$ is **observationally complete** with respect to $\Sigma_{\models}$ if $\Sigma_{\models} \subseteq \mathcal{O}(\Sigma)$* [16].

Intuitively, the definition of observational completeness can be interpreted as follows: $\Sigma$ is a representation of information (such as a single diagram or a set of set-theoretic sentences) and $\Sigma_{\models}$ is a set of statements whose truth we wish to establish. If we can simply observe those statements to be true from $\Sigma$ then $\Sigma$ is observationally complete with respect to $\Sigma_{\models}$.

**Definition 3.** *Let $\Sigma$ and $\hat{\Sigma}$ be finite, semantically equivalent sets of statements. Let $\sigma$ be a statement. If $\sigma$ is not observable from $\Sigma$ and $\sigma$ is observable from $\hat{\Sigma}$ then $\sigma$ is an **observational advantage** of $\hat{\Sigma}$ given $\Sigma$* [16].

---

[2] It is possible to define observability for other types of diagrams and statements too.

Using the requirement that observable statements must be semantically entailed, we see that any statement, $\sigma$, which is an observational advantage of $\hat{\Sigma}$ given $\Sigma$ is semantically entailed by $\Sigma$.

## 3 Set Theory and Euler Diagrams with Existential Import

To develop the theory of observation and observational advantages in the case of set theory and Euler diagrams with existential import, we require a formalisation of both systems. Since a ready comparison of statements needs to be made across notations, the set of labels used to denote sets will be common to both set theory and Euler diagrams, as will their interpretation.

**Definition 4.** *Define $\mathcal{L}$ to be a set whose elements are called **labels**. Two special symbols, $\emptyset$ and $U$, are not in $\mathcal{L}$* [16].

**Definition 5.** *An **interpretation** is a pair, $\mathcal{I} = (\triangle, \Psi)$, where $\triangle$ is a set and $\Psi$ is a function, $\Psi\colon \mathcal{L} \cup \{\emptyset, U\} \to \mathbb{P}\triangle$, that maps labels to subsets of $\triangle$ and ensures that $\Psi(\emptyset) = \emptyset$ and $\Psi(U) = \triangle$* [16].

### 3.1 Euler Diagrams with Existential Import

We now introduce the syntax and semantics of Euler diagrams with existential import. The syntax remains unchanged from [16] and is included here for ease of reference. The semantics, however, differ due to the requirement that zones must represent non-empty sets. To begin, we formally define zones and regions.

**Definition 6.** *A **zone** is a pair of finite, disjoint sets of labels, $(L_i, L_o)$, drawn from $\mathcal{L}$. A finite set of zones is a **region**.*

In Fig. 2, there are four zones. The zone inside just $P$ is $(\{P\}, \{Q, R, S\})$ and the zone outside all the curves is $(\emptyset, \{P, Q, R, S\})$. This diagram uses four labels, so we write $L = \{P, Q, R, S\}$, where $L$ denotes the diagram's label set. The diagram's set of zones will be denoted $Z$, so in this case

$$Z = \{(\{P\}, \{Q, R, S\}), (\{P, Q\}, \{R, S\}), (\{R, S\}, \{P, Q\}), (\emptyset, \{P, Q, R, S\})\}.$$

Formally, an Euler diagram is a set of labels together with a set of zones:

**Definition 7.** *An **Euler diagram**, $d$, is a pair, $(L, Z)$, where $L$ is a finite subset of $\mathcal{L}$, and for all zones, $(L_i, L_o)$, in $Z$ it is the case that $L_i \cup L_o = L$. Given $d = (L, Z)$, we sometimes write $L(d)$ and $Z(d)$ for $L$ and $Z$ respectively. Given a finite set, $\mathcal{D}$, of Euler diagrams we define $L(\mathcal{D})$ to be $\bigcup\limits_{d \in \mathcal{D}} L(d)$.*

To define the semantics of Euler diagrams, it is useful to identify the zones that *could* be present in the diagram given the labels used, but which are in fact *missing*. Intuitively, missing zones represent the empty set.

**Definition 8.** *Let $d = (L, Z)$ be an Euler diagram. The* ***missing zones*** *of $d$ are elements of $MZ(d) = \{(L_i, L \backslash L_i) : L_i \subseteq L\} \backslash Z$.*

We now extend the definition of an interpretation to identify the sets represented by zones and regions:

**Definition 9.** *Let $\mathcal{I} = (\triangle, \Psi)$ be an interpretation. An extension of $\Psi$ to map zones and regions to sets is defined as follows:*

1. *for each zone, $(L_i, L_o)$, $\Psi(L_i, L_o) = \bigcap_{l \in L_i} \Psi(l) \cap \bigcap_{l \in L_o} \overline{\Psi(l)}$, and*
2. *for each region, $r$, $\Psi(r) = \bigcup_{(L_i, L_o) \in r} \Psi(L_i, L_o)$.*

Our next task is to define the circumstances under which an interpretation is a model for (i.e., agrees with the intuitive meaning of) an Euler diagram. As well as missing zones representing empty sets we also have to account for the existential import requirement: present zones represent non-empty sets:

**Definition 10.** *Let $d = (L, Z)$ be an Euler diagram and $\mathcal{I} = (\triangle, \Psi)$ be an interpretation. Then $\mathcal{I}$* ***satisfies*** *$d$ and is a* ***model*** *for $d$ whenever $\Psi(z) \neq \emptyset$ for each zone $z$ in $Z$ and $\Psi(z) = \emptyset$ for each zone $z$ in $MZ(d)$.*

### 3.2 Set-Theoretic Sentences

We now extend the work in [16] on set-theoretic sentences to allow statements to be made with $\neq$ and $\not\subseteq$, as well as $=$ and $\subseteq$. Firstly, we define *set-theoretic expressions*, which are syntactic representations of sets formed from the 'basic sets' represented by labels in $\mathcal{L}$:

**Definition 11.** *The following are* ***set-theoretic expressions*** *or, simply, set-expressions: (i) $U$ and $\emptyset$ are both set-expressions, (ii) every label in $\mathcal{L}$ is a set-expression, and (iii) if $s_1$ and $s_2$ are set-expressions then so are $(s_1 \cap s_2)$, $(s_1 \cup s_2)$, $(s_1 \backslash s_2)$, and $\overline{s_1}$* [16].

Given labels $P$, $Q$ and $R$, the following are some examples of set-theoretic expressions (omitting unnecessary brackets): $P$, $P \cap Q$, $Q \cup R$, $P \backslash (Q \cup R)$ and $\overline{(P \cap Q)}$. Often we will blur the distinction between syntax and semantics, talking of 'the set $P \cap Q$' when strictly speaking we mean the set represented by $P \cap Q$; given an interpretation, $(\triangle, \Psi)$, this set is $\Psi(P) \cap \Psi(Q)$. Set-theoretic expressions merely construct sets from the basic ones. We can then make assertions about the relationship between set-theoretic expressions using $\subseteq$, $\not\subseteq$, $=$, and $\neq$:

**Definition 12.** *Given set-expressions $s_1$ and $s_2$ the following are* ***set-theoretic sentences:*** *$s_1 \subseteq s_2$, $s_1 \not\subseteq s_2$, $s_1 = s_2$, and $s_1 \neq s_2$. Sentences of the form $s_1 \subseteq s_2$ and $s_1 = s_2$ are* ***positive*** *whereas those of the form $s_1 \not\subseteq s_2$ and $s_1 \neq s_2$ are* ***negative****.*

When we want to give set-expressions or set-theoretic sentences names, we will use $\equiv$. For example, to refer to $P \cap Q$ and $R \subseteq P \cap Q$ by the names $s_1$ and $s_2$, we write $s_1 \equiv P \cap Q$ and $s_2 \equiv R \subseteq P \cap Q$. This is to avoid overloading $=$. It is also helpful to us to have access to set of the labels, denoted $L(s)$, used in any given set-theoretic sentence, $s$: $L(s)$ is defined in the obvious recursive way. We extend this to a finite set, $\mathcal{S}$, of set-theoretic sentences: $L(\mathcal{S})$ is $\bigcup\limits_{s \in \mathcal{S}} L(s)$, that is, the set of all labels appearing in members of $\mathcal{S}$.

Now, to reiterate, every set-theoretic sentence only has one meaning-carrier: the set-expression on the left is in the asserted relationship with the set-expression on the right. This leads us to the definition of the semantics of set-theoretic sentences. The labels over which set-expressions are formed are already interpreted as sets (Definition 5). We extend this to cover the interpretation of more complex set-expressions in order to identify when an interpretation 'agrees with' the intuitive meaning of (i.e., is a model for) sentences.

**Definition 13.** *Let $\mathcal{I} = (\triangle, \Psi)$ be an interpretation. An extension of $\Psi$ to map set-expressions to sets is defined as follows. For each set-expression, $s$,*

1. *if $s \in \mathcal{L} \cup \{U, \emptyset\}$ then $\Psi(s)$ is already defined,*
2. *if $s \equiv (s_1 \star s_2)$, where $\star \in \{\cap, \cup, \backslash\}$, then $\Psi(s) = \Psi(s_1) \star \Psi(s_2)$, and*
3. *if $s \equiv \overline{s_1}$ then $\Psi(s) = \overline{\Psi(s_1)} = \Psi(U) \backslash \Psi(s_1)$.*

**Definition 14.** *Let $s$ be a set-theoretic sentence. Let $\mathcal{I} = (\triangle, \Psi)$ be an interpretation. Then $\mathcal{I}$* ***satisfies*** *$s$ and is a* ***model*** *for $s$ under the following circumstances:*

1. *if $s \equiv s_1 \subseteq s_2$ then $\Psi(s_1) \subseteq \Psi(s_2)$,*
2. *if $s \equiv s_1 \not\subseteq s_2$ then $\Psi(s_1) \not\subseteq \Psi(s_2)$,*
3. *if $s \equiv s_1 = s_2$ then $\Psi(s_1) = \Psi(s_2)$, and*
4. *if $s \equiv s_1 \neq s_2$ then $\Psi(s_1) \neq \Psi(s_2)$.*

*Let $\mathcal{S}$ be a finite set of set-theoretic sentences. Then $\mathcal{I}$* ***satisfies*** *$\mathcal{S}$ and is a* ***model*** *for $\mathcal{S}$ provided $\mathcal{I}$ is a model for each set-theoretic sentence in $\mathcal{S}$.*

### 3.3 Semantic Relationships

The final prerequisite for studying the observational advantages of Euler diagrams over set-theoretic sentences relies on us tying up their semantic relationships, beyond just mapping their (common) labels to sets in interpretations. We generically refer to Euler diagrams and set-theoretic sentences as *statements*.

**Definition 15.** *Let $\sigma_1$ and $\sigma_2$ be statements. If $\sigma_1$ and $\sigma_2$ have the same models then they are* ***semantically equivalent****. If finite sets of statements, $\Sigma_1$ and $\Sigma_2$, have the same models then they are* ***semantically equivalent****.*

**Definition 16.** *Let $\Sigma$ be a finite set of statements and let $\sigma$ be a statement. Then $\Sigma$* ***semantically entails*** *$\sigma$, denoted $\Sigma \vDash \sigma$, provided every model for $\Sigma$ is also a model for $\sigma$. If $\sigma$ is semantically entailed by, but not in $\Sigma$, then $\sigma$ is* ***properly*** *semantically entailed by $\Sigma$.*

Lastly, since our focus is on the observational completeness of Euler diagrams with respect to set-theoretic sentences and the conditions under which this can be achieved, it is useful for us to introduce notation for the set of all set-theoretic sentences that are properly entailed given the labels used:

**Definition 17.** *Let $\mathcal{S}$ be a finite set of set-theoretic sentences. Define $\mathcal{S}_{\models}^{L(\mathcal{S})}$ to be the set of set-theoretic sentences that are properly semantically entailed by $\mathcal{S}$ such that each $s \in \mathcal{S}_{\models}^{L(\mathcal{S})}$ ensures $L(s) \subseteq L(\mathcal{S})$.*

We can think of the labels used – that is, those in $L(\mathcal{S})$ – as being the sets of interest, since these are the sets about which $\mathcal{S}$ provides information. Then we can view $\mathcal{S}_{\models}^{L(\mathcal{S})}$ as containing precisely the set-theoretic sentences that make true statements about the sets of interest, but which are not explicitly given in $\mathcal{S}$. In other words, these are the statements that we can and must *infer* from $\mathcal{S}$.

## 4 Model Theory: Euler Diagrams with Existential Import

A major consideration for us is to identify when, given a set of set-theoretic sentences, $\mathcal{S}$, there exists a semantically equivalent Euler diagram, $d$. This is a prerequisite for identifying whether $d$ is observationally complete with respect to $\mathcal{S}$. Our strategy for this is to provide insight into what the models of Euler diagrams 'look like'. Unsurprisingly, this section establishes that the models for Euler diagrams with existential import are those for which all of the present zones represent non-empty sets and the missing zones represent empty sets. As it will be beneficial to us later, we define a relation on interpretations inspired by this insight:

**Definition 18.** *Let $L \subseteq \mathcal{L}$ be a set of labels and $\mathcal{I}_1 = (\triangle_1, \Psi_1)$ and $\mathcal{I}_2 = (\triangle_2, \Psi_2)$ be interpretations. Then $\mathcal{I}_1$ and $\mathcal{I}_2$ are **L-approximate**, denoted $\mathcal{I}_1 \approx_L \mathcal{I}_2$, provided for every zone $(L_i, L_o)$ where $L_i \cup L_o = L$, $\Psi_1(L_i, L_o) = \emptyset$ iff $\Psi_2(L_i, L_o) = \emptyset$.*

Intuitively, two interpretations are $L$-approximate if one never assigns the empty set to a zone formed over $L$ when the other does not. Clearly, $\approx_L$ is an equivalence relation on the set of interpretations.

**Theorem 1.** *Let $d = (L, Z)$ be an Euler diagram. Then the set of models, $M(d)$, for $d$ is an equivalence class of interpretations under $\approx_L$. In particular, $\mathcal{I} = (\triangle, \Psi)$ is in $M(d)$ iff for each zone, $z_p$, in $Z(d)$, we have $\Psi(z_p) \neq \emptyset$ and for each zone, $z_m$, in $MZ(d)$ we have $\Psi(z_p) = \emptyset$.*

Theorem 1 demonstrates the highly constrained nature of models for Euler diagrams with existential import: they are single equivalence classes of $L$-approximate interpretations, forcing present zones to represent non-empty sets. There is no possibility for representing uncertainty when it comes to the non-emptiness of a set. By contrast, if the existential import requirement is removed (so a present zone can be empty or not) then the model sets are *unions* of equivalence classes: the models are given by $M_{Z_1} \cup ... \cup M_{Z_{2^n}}$ where: $Z_1$ to $Z_{2^n}$ are the $2^n$ subsets of $Z$, and $M_{Z_i}$ is the equivalence class of interpretations where for each zone, $z_p$ in $Z_i$, we have $\Psi(z_p) \neq \emptyset$ and for each zone, $z_m$ in $(Z(d) \backslash Z_i) \cup MZ(d)$ we have $\Psi(z_p) = \emptyset$.

## 5 Model Theory: Set-Theoretic Sentences

In order to identify when a set of set-theoretic sentences, $\mathcal{S}$, has a semantically equivalent diagram, $d$, we start by appealing to Theorem 1. This theorem tells us that $\mathcal{S}$ *only* has such a diagram if its models are also a single equivalence class under $\approx_L$. Clearly, such an equivalence class determines a set of zones that represent (non)empty sets. It is therefore useful to introduce the idea of *determining set-emptiness*:

**Definition 19.** *Let $\mathcal{S}$ be a finite set of set-theoretic sentences. We say $\mathcal{S}$* ***determines set-emptiness*** *if the set of models, $M(\mathcal{S})$, for $\mathcal{S}$ forms an equivalence class of interpretations under $\approx_{L(\mathcal{S})}$.*

For example, consider the following:

$\mathcal{S}_1 = \{P \not\subseteq Q, Q \not\subseteq P\}$,
$\mathcal{S}_2 = \{P \not\subseteq Q, Q \not\subseteq P, P \cap Q = \emptyset\}$, and
$\mathcal{S}_3 = \{P \not\subseteq Q, Q \not\subseteq P, P \cap Q = \emptyset, \overline{P \cup Q} \neq \emptyset\}$.

Note $L(\mathcal{S}_1) = L(\mathcal{S}_2) = L(\mathcal{S}_3) = \{P, Q\}$. Among these, only $\mathcal{S}_3$ determines set-emptiness. $\mathcal{S}_1$ does not, since it can be satisfied by both, an interpretation assigning the empty set to zone $(\{P, Q\}, \{\})$, and one assigning a non-empty set to it, for example. These interpretations are not $L(\mathcal{S}_1)$-approximate, so $\mathcal{S}_1$ does not determine set-emptiness. With the addition of sentence $P \cap Q = \emptyset$, the set of models for $\mathcal{S}_2$ no longer has interpretations that 'disagree' on the zone $(\{P, Q\}, \emptyset)$ (every model assigns the empty set to it), yet models can still disagree on zone $(\emptyset, \{P, Q\})$. Adding another sentence, $\overline{P \cup Q} \neq \emptyset$, to give $\mathcal{S}_3$ makes all models agree on the zones that can be formed over $\{P, Q\}$: they assign the empty set to $(\emptyset, \{P, Q\})$ and non-empty sets to $(\{P\}, \{Q\})$, $(\{Q\}, \{P\})$, and $(\emptyset, \{P, Q\}))$; a semantically equivalent Euler diagram is given in Fig. 3. Thus, determining set-emptiness is rather a high demand to place on the case of set-theoretic sentences: only very limited sets of set-theoretic sentences determine set-emptiness. By contrast, determining set-emptiness is not placing such a high demand on Euler diagrams. Indeed, every single Euler diagram determines set-emptiness. We obtain the following lemma:

**Lemma 1.** *Let $\mathcal{S}$ be a set of set-theoretic sentences. Then $\mathcal{S}$ has a semantically equivalent Euler diagram only if $\mathcal{S}$ determines set-emptiness.*

This lemma indicates the extent of the overspecificity of Euler diagrams, relative to set-theoretic sentences. The phenomenon of overspecificity has been pointed out in [17] and further investigated in [14] in connection to a wider variety of diagrams. However, the impact of the phenomenon in a specific diagrammatic system has never been formalised. Our approach illustrates how it can be investigated.

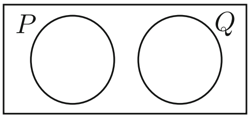


**Fig. 3.** Translating set-theoretic sentences.

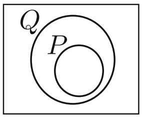


**Fig. 4.** Relevant zones.

Our next goal is to characterise the sets of set-theoretic sentences that meet the demand of determining set-emptiness. To begin, we notice that *positive* set-theoretic sentences provide information about the emptiness of sets; $P \subseteq Q$ tells us that $P \backslash Q = \emptyset$ and $R = S$ tells us that $R \backslash S = \emptyset$ and $S \backslash R = \emptyset$. Moreover, *negative* set-theoretic sentences provide information about the non-emptiness of sets; $P \nsubseteq Q$ expresses $P \backslash Q \neq \emptyset$ and $R \neq S$ implies $R \backslash S \neq \emptyset$ or $S \backslash R \neq \emptyset$. It is therefore useful to distinguish the positive and negative cases:

**Definition 20.** *Given a set of set-theoretic sentences $\mathcal{S}$, we define $\mathcal{S}^+$ and $\mathcal{S}^-$ to be the set of all positive and negative members of $\mathcal{S}$, respectively.*

So, positive sentences provide information about empty zones whereas negative sentences provide information about non-empty zones. This leads to the idea of a *relevant zone*, which relies on a translation of a set-theoretic sentence to a region which is determined by the sets of interest. For example, given $L = \{P, Q\}$ as the sets of interest, the expression $P \backslash Q$ corresponds to the zone $(\{P\}, \{Q\})$ since, informally, $(\{P\}, \{Q\})$ represents the set of things in $P$ that are not in $Q$, that is, $P \backslash Q$. Likewise, the expression $P$ – again given $L = \{P, Q\}$ – corresponds to the region $\{(\{P\}, \{Q\}), (\{P, Q\}, \emptyset)\}$: the elements in $P$ can be either in $P \backslash Q$, corresponding to $(\{P\}, \{Q\})$, or in $P \cap Q$, corresponding to $(\{P, Q\}, \emptyset)$.

**Definition 21.** *Let $s$ be a set-expression and let $L$ be a set of labels such that $L(s) \subseteq L$. The* ***translation*** *of $s$ given $L$ into a region, denoted $\mathcal{T}(s, L)$, is defined recursively:*

1. *if $s \equiv \emptyset$ then $\mathcal{T}(s, L) = \emptyset$,*
2. *if $s \equiv U$ then $\mathcal{T}(s, L) = \{(L_i, L_o) : L_i \cup L_o = L \wedge L_i \cap L_o = \emptyset\}$,*
3. *if $s \in \mathcal{L}$ then $\mathcal{T}(s, L) = \{(L_i, L_o) \in \mathcal{T}(U, L) : s \in L_i\}$,*
4. *if $s \equiv (s_1 \star s_2)$, where $\star \in \{\cap, \cup, \backslash\}$, then $\mathcal{T}(s, L) = (\mathcal{T}(s_1, L) \star \mathcal{T}(s_2, L))$, and*
5. *if $s \equiv \overline{s_1}$ then $\mathcal{T}(s, L) = (\mathcal{T}(U, L) \backslash \mathcal{T}(s_1, L))$.*

Using the translation of set-expressions to regions, we can now see how to translate set-theoretic sentences to regions too. For instance, $P \subseteq Q$ is true whenever $\Psi(P) \subseteq \Psi(Q)$. In terms of zones formed over $P$ and $Q$, the sentence $P \subseteq Q$ is true whenever

$$\Psi(\{(\{P\},\{Q\}),(\{P,Q\},\emptyset)\}) \subseteq \Psi((\{Q\},\{P\}),(\{P,Q\},\emptyset)). \qquad (*)$$

Figure 4 illustrates $P \subseteq Q$ and we see that the zone $(\{P\},\{Q\})$ is missing. Therefore (*) is true, and the zone $(\{P\},\{Q\})$ is *relevant* in this case.

**Definition 22.** *Given a set-theoretic sentence $s$ and a set of labels $L$ such that $L(s) \subseteq L$, we define the* ***relevant set of zones of s given L****, denoted $\mathcal{RZ}(s,L)$, in the following way:*

1. *If $s$ is of the form $s_1 = s_2$ or $s_1 \neq s_2$, $\mathcal{RZ}(s,L) = (\mathcal{T}(s_1,L)\backslash\mathcal{T}(s_2,L)) \cup (\mathcal{T}(s_2,L)\backslash\mathcal{T}(s_1,L))$,*
2. *If $s$ is of the form $s_1 \subseteq s_2$ or $s_1 \not\subseteq s_2$, $\mathcal{RZ}(s,L) = \mathcal{T}(s_1,L)\backslash\mathcal{T}(s_2,L))$.*

So, continuing with the example above, we have

$$\mathcal{RZ}(P \subseteq Q,\{P,Q\}) = \{(\{P\},\{Q\})\}$$

and, whenever $P \subseteq Q$, we know that $(\{P\},\{Q\})$ represents the empty set. So, the relevant set of zones of a set-theoretic sentence, $s$, is 'relevant' to $s$ in that the zones help to determine when $s$ is satisfied by an interpretation. The following lemma makes this point more precise.

**Lemma 2.** *Let $s$ be a set-theoretic sentence and $\mathcal{I} = (\triangle,\Psi)$ be an interpretation. Let $L$ be a set of labels such that $L(s) \subseteq L$. Then*

1. *if $s$ is positive then $\mathcal{I}$ is a model for $s$ iff $\Psi(\mathcal{RZ}(s,L)) = \emptyset$,*
2. *if $s$ is negative then $\mathcal{I}$ is a model for $s$ iff $\Psi(\mathcal{RZ}(s,L)) \neq \emptyset$ for some zone $z \in \mathcal{RZ}(s,L)$.*

Our next goal is to identify conditions under which any set of set-theoretic sentences, $\mathcal{S}$, determines set-emptiness. To produce such conditions, it is important to have an understanding of what the models for $\mathcal{S}$ 'look like'. We can gain such insight by considering the models for $\mathcal{S}^+$ and $\mathcal{S}^-$ separately, informed by Lemma 2, noting that the models for $\mathcal{S}$ must model both $\mathcal{S}^-$ and $\mathcal{S}^+$.

The set of relevant zones, in the case of positive set-theoretic sentences, gives us information about which zones *must* represent the empty set. In this sense, the positive set-theoretic sentences in $\mathcal{S}$ partially characterise the models for $\mathcal{S}$. By Lemma 2, an interpretation, $\mathcal{I} = (\triangle,\Psi)$, is in $M(\mathcal{S}^+)$ (the set of models for $\mathcal{S}^+$) iff, for each $s$ in $\mathcal{S}^+$, $\Psi(\mathcal{RZ}(s,L(\mathcal{S})) = \emptyset$. Therefore, $\mathcal{I} = (\triangle,\Psi)$ is in $M(\mathcal{S}^+)$ provided

$$\bigcup_{s\in\mathcal{S}^+} \Psi(\mathcal{RZ}(s,L(\mathcal{S})) = \emptyset.$$

For ease of notation, we define the **empty zones** of $\mathcal{S}$ to be elements of

$$\mathcal{EZ}(\mathcal{S}) = \bigcup_{s\in\mathcal{S}^+} \mathcal{RZ}(s,L(\mathcal{S}))$$

and the **Venn zones** of $\mathcal{S}$ to be elements of

$$\mathcal{VZ}(\mathcal{S}) = \{(L_i, L(\mathcal{S})\backslash L_i) : L_i \subseteq L(\mathcal{S})\}.$$

The empty zones represent empty sets in all models for $\mathcal{S}$. The remaining zones in the Venn zone set *may or may not* represent empty sets. Now, we have some information about non-emptiness, provided by $\mathcal{S}^-$, but it need not completely determine whether any given zone is *necessarily* non-empty in a model. This is where $\mathcal{S}^-$ must be considered carefully.

From Lemma 2, we know that an interpretation, $\mathcal{I} = (\triangle, \Psi)$, is in $M(\mathcal{S}^-)$ iff, for each $s^-$ in $\mathcal{S}^-$, $\Psi(z) \neq \emptyset$ for some $z \in \mathcal{RZ}(s^-, L(\mathcal{S}))$. For $\mathcal{S}$ to determine set-emptiness, therefore, we seek conditions on $\mathcal{S}^-$ that are necessary and sufficient to ensure that *each* zone in $\mathcal{VZ}(\mathcal{S})\backslash\mathcal{EZ}(\mathcal{S})$ represents a non-empty set.

In this context, we aim to identify sets of zones that partially characterise some of the models for $\mathcal{S}^-$: given a set of zones, $Z$, under what conditions is the set of interpretations that map the zones in $Z$ to *non-empty sets* a set of models for $\mathcal{S}^-$? As a first step, we introduce the idea of a *choice function*, which assigns relevant zones to negative set-theoretic sentences. Importantly, assigned zones cannot be empty zones.

**Definition 23.** *Let $\mathcal{S}$ be a finite set of set-theoretic sentences. A **choice function**, $c\colon \mathcal{S}^- \to \mathcal{VZ}(\mathcal{S})\backslash\mathcal{EZ}(\mathcal{S})$ for $\mathcal{S}$, maps negative set-theoretic sentences in $\mathcal{S}$ to zones such that for each $s^- \in \mathcal{S}^-$, $c(s^-) \in \mathcal{RZ}(s^-, L(\mathcal{S}))$.*

Clearly, given an arbitrary $\mathcal{S}$ there need not exist a choice function. This occurs when there is a negative set-theoretic sentence in $\mathcal{S}$ such that all of its relevant zones are in $\mathcal{EZ}(\mathcal{S})$. Under such circumstances, it is obvious that $\mathcal{S}$ has no models and is, therefore, inconsistent. However, given an arbitrary choice function, $c$, the zones in $\mathcal{VZ}(\mathcal{S})\backslash\mathcal{EZ}(\mathcal{S})$ to which $c$ maps set-theoretic sentences (i.e., the set of zones that is the image of $c$) partially characterises some of the models for $\mathcal{S}^-$: all interpretations where these zones represent non-empty sets are models for $\mathcal{S}^-$. Intuitively, any given model for $\mathcal{S}^-$ is classified by some choice function.

So far, we have characterised all of the models for $\mathcal{S}^+$ and the models for $\mathcal{S}^-$. In a build-up to our set of necessary and sufficient conditions that identify when $\mathcal{S}$ defines set-emptiness, we establish when $\mathcal{S}$ is satisfiable, using choice functions. We start by building an interpretation using a choice function.

**Definition 24.** *Let $\mathcal{S}$ be a finite set of set-theoretic sentences for which there exists a choice function, $c\colon \mathcal{S}^- \to \mathcal{VZ}(\mathcal{S})\backslash\mathcal{EZ}(\mathcal{S})$. We define the **choice interpretation** for $\mathcal{S}$ given $c$ to be the interpretation $\mathcal{I}^c_{\mathcal{S}} = (\triangle, \Psi)$ as follows:*

1. *the universal set, $\triangle$, is the image of $c$, that is:*

$$\triangle = \{z \in \mathcal{VZ}(\mathcal{S})\backslash\mathcal{EZ}(\mathcal{S}) : \exists s^- \in \mathcal{S}^- \; c(s^-) = z\}, \qquad \text{and}$$

2. *for each $l \in \mathcal{L}$, we define*

$$\Psi(l) = \begin{cases} \{(L_i, L_o) \in \triangle : l \in L_i\} & \textit{if } l \in L(\mathcal{S}) \\ \emptyset & \textit{otherwise.} \end{cases}$$

Lemma 3 establishes that the choice interpretation is a model for $\mathcal{S}$:

**Lemma 3.** *Let $\mathcal{S}$ be a finite set of set-theoretic sentences for which there exists a choice function, $c\colon \mathcal{S}^- \to \mathcal{VZ}(\mathcal{S})\backslash\mathcal{EZ}(\mathcal{S})$. The choice interpretation, $\mathcal{I}^c_{\mathcal{S}} = (\triangle, \Psi)$, for $\mathcal{S}$ given $c$ is a model for $\mathcal{S}$.*

Lemma 3 builds on our insight into what sets of models 'look like' for $\mathcal{S}$. We have seen that choice functions can be used to define models. Importantly, the absence of a choice function implies the absence of models: $\mathcal{S}$ is unsatisfiable.

Choice functions with different images correspond to models that are not $L(\mathcal{S})$-approximate. In particular, if there is a non-surjective choice function then there are necessarily models for $\mathcal{S}$ that are not $L(\mathcal{S})$-approximate. This semantic intuition is captured syntactically via choice functions in Theorem 2.

**Theorem 2.** *Let $\mathcal{S}$ be a finite set of set-theoretic sentences. Then $\mathcal{S}$ determines set-emptiness iff there exists a choice function for $\mathcal{S}$ and all choice functions for $\mathcal{S}$ are surjective.*

Thus, Theorem 2 is what is needed to meet our major goal for this section: the provision of necessary and sufficient conditions for determining set-emptiness. The models for such an $\mathcal{S}$ are characterised by the following theorem:

**Theorem 3.** *Let $\mathcal{S}$ be a finite set of set-theoretic sentences that determines set-emptiness. Let $\mathcal{I} = (\triangle, \Psi)$ be an interpretation. Then $\mathcal{I}$ is a model for $\mathcal{S}$ iff*

1. *the empty zones of $\mathcal{S}$ all represent the empty set: $\Psi(\mathcal{EZ}(\mathcal{S})) = \emptyset$, and*
2. *the remaining zones all represent non-empty sets: for all $z \in \mathcal{VZ}(\mathcal{S})\backslash\mathcal{EZ}(\mathcal{S})$, $\Psi(z) \neq \emptyset$.*

## 6 Observational Completeness

We now set out to identify an Euler diagram that is observationally complete given a set-emptiness defining $\mathcal{S}$. Focusing first on the requisite Euler diagram, $d$, for $\mathcal{S}$, we need $d$ to have the same models as $S$. That is, $d$'s present zones (which represent non-empty sets) should correspond to $\mathcal{VZ}(\mathcal{S})\backslash\mathcal{EZ}(\mathcal{S})$, since these are precisely the zones that represent non-empty sets in all models for $\mathcal{S}$. Likewise, the zones not in $d$ should correspond to those in $\mathcal{EZ}(\mathcal{S})$, since these represent empty sets in all models for $\mathcal{S}$. We have already seen an example of the Euler diagram for a given set of set-theoretic sentences in Fig. 3, given $\mathcal{S}_3$ on page 9.

**Definition 25.** *Let $\mathcal{S}$ be a finite set of set-theoretic sentences that determines set-emptiness. The **Euler diagram for $\mathcal{S}$**, denoted $d_{\mathcal{S}}$, is*

$$d_{\mathcal{S}} = (L(\mathcal{S}), \mathcal{VZ}(\mathcal{S})\backslash\mathcal{EZ}(\mathcal{S})).$$

Importantly, $\mathcal{S}$ and $d_{\mathcal{S}}$ are semantically equivalent, which follows from Theorems 1 and 3:

**Theorem 4.** *Let $\mathcal{S}$ be a finite set of set-theoretic sentences that determines set-emptiness. Then $\mathcal{S}$ and $d_{\mathcal{S}}$ are semantically equivalent.*

We must now consider what it means for a sentence to be observable from an Euler diagram, generalising [16]. To do this, we need to translate *regions* to set-expressions. Intuitively, regions translate to multiple set-expressions. For instance, in Fig. 4, the region comprising the single zone inside the curve $P$ corresponds to various set-expressions, including $P$ and $P \cap Q$, since this zone represents both the set $P$ and the set $P \cap Q$; indeed, in this case $P = P \cap Q$. For our purposes here it is sufficient to have an intuitive understanding of what set-expressions can arise from regions, along the lines of the example just given[3]. Using this intuitive approach, we can now define observability:

**Definition 26.** *Let $d$ be an Euler diagram and let $s \equiv s_1 \star s_2$, where $\star \in \{\subseteq, \not\subseteq, =, \neq\}$, be a set-theoretic sentence. Then $s \equiv s_1 \star s_2$ is* ***observable*** *from $d$ provided there exist regions $r_1$ and $r_2$ of $d$ such that*

1. *$r_1 \star r_2$,*
2. *$s_1$ is a translation of $r_1$, and*
3. *$s_2$ is a translation of $r_2$.*

Finally, we have one of our key results:

**Theorem 5.** *Let $\mathcal{S}$ be a finite set of set-theoretic sentences that determines set-emptiness. Then $\{d_{\mathcal{S}}\}$ is observationally complete with respect to $\mathcal{S}_{\models}^{L(\mathcal{S})}$.*

## 7 Discussion

Our results on Euler diagrams with existential import demonstrate that there are severe limitations due to overspecificity, at least from the perspective of observational advantages. This is potentially problematic since diagrams, by their very nature, are believed to excel as representations of information due to their ability to make facts explicit that would otherwise need to be inferred.

To recap, an Euler diagram, $d$, with existential import is only semantically equivalent to a finite set, $\mathcal{S}$, of set-theoretic sentences when $\mathcal{S}$ determines set-emptiness. This is a serious limitation, arising because the models for $d$ are a single equivalence class under the $L$-approximate relation. The crux of the problem is that such diagrams require complete certainty over whether zones represent empty sets. By contrast, most sets of set-theoretic sentences do not make this demand on their model sets and are, in this case, more expressive than their diagrammatic counterpart.

This suggests that diagrams which allow uncertainty to be expressed, and thus avoid overspecificity, are more likely to have observational advantages over competing notations. Indeed, suppose that the existential import requirement is

---

[3] It is straightforward, yet lengthy, to define a translation from regions to set-expressions; due to space constraints, we refer the reader to [16].

removed and, instead, Peirce's $\otimes$-sequences are used to express non-emptiness. We conjecture that any finite set, $\mathcal{S}$, of set-theoretic sentences (as in Definition 12) will be semantically equivalent to some diagram, $d$. Moreover, we expect $d$ to be observationally complete with respect to $\mathcal{S}_{\models}^{L(\mathcal{S})}$. Such diagrams do not suffer from overspecificity issues and have models that are unions of equivalence classes under the $L$-approximate relation, just like sets of set-theoretic sentences.

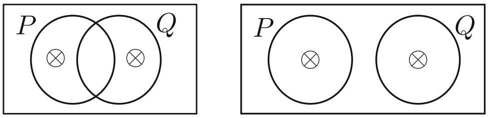


**Fig. 5.** Exploiting $\otimes$-sequences to overcome overspecificity limitations.

To illustrate, consider again $\mathcal{S}_1 = \{P \nsubseteq Q, Q \nsubseteq P\}$. Whilst no Euler diagram with existential import can express this information, we could use an Euler diagram with $\otimes$-sequences instead to define non-emptiness (left of Fig. 5). From the diagram, we can observe, for instance, that $P \backslash Q \neq \emptyset$ due to $\otimes$ inside $P$ but outside $Q$. The diagram on the right of Fig. 5 illustrates how we can depict $\mathcal{S}_2 = \{P \nsubseteq Q, Q \nsubseteq P, P \cap Q = \emptyset\}$. It will be interesting to extend the work in this paper to determine whether this alternative system of Euler diagrams is indeed observationally complete for any given $\mathcal{S}$. Importantly, in this alternative system, the zones containing no $\otimes$ symbol can represent either empty or non-empty sets, thus removing the overspecificity arising from existential import.

## 8 Conclusion

The ideas of observation, observational advantages and observational completeness enable us to formally compare different representations of information. It is considered *advantageous* if a representation of information simply allows us to observe other statements of interest to be true. Therefore, this suggests that designing notations that allow many observations to be made, especially compared to competing representations, is sensible. In the case of diagrams, free rides and observational advantages are seen as a major feature that indicates how and when they *may* be more efficacious than symbolic or textual notations.

We demonstrated that overspecificity makes diagrams less observationally advantageous. As in the case of Euler diagrams with existential import, overspecificity often means *there is no corresponding diagram for a given representation of information.* This is clearly undesirable and leads us to posit that diagrams should be carefully designed in order to ensure that they do not have overspecifity issues and also support the observability of information. Indeed, our results indicate an advantage of set-theoretic language: it can express information freely, whether the information is strong enough to determine set-emptiness or not. Euler diagrams with existential import are disadvantageous in that respect.

There is still much work to be done, however, to ascertain the extent to which observational advantages are also cognitive advantages. We think it is important to understand the *net cognitive value* of observability. There is certainly cognitive cost associated with observing statements, but to what extent is this cost 'lower' than the alternative task of *inferring* information instead? The net cognitive value of a statement observable from a diagram depends on the cost of recognising a meaning-carrying relationship and also on the set of available operations to translate this meaning-carrier into an alternative representation. This research needs to be, in the future, connected to a psychological and computational model of the perceptual operations available to people alongside the formal investigations that we have begun. Preliminary work in [1] is exploring this important cognitive aspect, and it will be interesting to see how it develops.

**Acknowledgements.** Stapleton and Jamnik were funded by a Leverhulme Trust Research Project Grant (RPG- 2016-082) for the project entitled Accessible Reasoning with Diagrams.

## References

1. Baigelenov, A., Saenz, M., Hung, Y.H., Parsons, P.: Toward an understanding of observational advantages in information visualization. In: IEEE Conference on Information Visualization, Poster Abstracts (2017)
2. Barwise, J., Etchemendy, J.: Hyperproof. CSLI Press, Stanford (1994)
3. Chatti, S., Schang, F.: The cube, the square and the problem of existential import. Hist. Philos. Logic **34**(2), 101–132 (2013)
4. Choudhury, L., Chakraborty, M.K.: On extending venn diagram by augmenting names of individuals. In: Blackwell, A.F., Marriott, K., Shimojima, A. (eds.) Diagrams 2004. LNCS (LNAI), vol. 2980, pp. 142–146. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-25931-2_14
5. Dretske, F.: Seeing and Knowing. Routledge & Kegan Paul, London (1969)
6. Gil, J., Howse, J., Kent, S.: Formalising spider diagrams. In: IEEE Symposium on Visual Languages, pp. 130–137. IEEE (1999)
7. Hammer, E., Shin, S.J.: Euler's visual logic. Hist. Philos. Logic **19**, 1–29 (1998)
8. Howse, J., Stapleton, G., Taylor, J.: Spider diagrams. LMS J. Comput. Math. **8**, 145–194 (2005)
9. Kent, S.: Constraint diagrams: visualizing invariants in object oriented models. In: Proceedings of OOPSLA 1997, pp. 327–341. ACM Press, October 1997
10. Mineshima, K., Okada, M., Takemura, R.: A diagrammatic inference system with Euler circles. J. Logic Lang. Inform. **21**(3), 365–391 (2012)
11. Moktefi, A., Pietarinen, A.V.: On the diagrammatic representation of existenial statements with Venn diagrams. J. Logic Lang. Inform. **24**(4), 361–374 (2015)
12. Peirce, C.: Collected Papers, vol. 4. Harvard University Press, Cambridge (1933)
13. Sato, Y., Mineshima, K., Takemura, R.: The efficacy of Euler and Venn diagrams in deductive reasoning: empirical findings. In: Goel, A.K., Jamnik, M., Narayanan, N.H. (eds.) Diagrams 2010. LNCS (LNAI), vol. 6170, pp. 6–22. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14600-8_6
14. Shimojima, A.: Semantic Properties of Diagrams and Their Cognitive Potentials. CSLI, Stanford (2015)

15. Shin, S.J.: The Logical Status of Diagrams. Cambridge University Press, Cambridge (1994)
16. Stapleton, G., Jamnik, M., Shimojima, A.: What makes an effective representation of information: a formal account of observational advantages. J. Logic Lang. Inform. **26**(2), 143–177 (2017)
17. Stenning, K., Oberlander, J.: A cognitive theory of graphical and linguistic reasoning: logic and implementation. Cogn. Sci. **19**(1), 97–140 (1995)
18. Swoboda, N., Allwein, G.: Using DAG transformations to verify Euler/Venn homogeneous and Euler/Venn FOL heterogeneous rules of inference. J. Softw. Syst. Model. **3**(2), 136–149 (2004)

# Observational Advantages: A Philosophical Discussion

Francesco Bellucci

Department of Philosophy and Communication, University of Bologna,
Via Azzo Gardino 23, 40122 Bologna, Italy
francesco.bellucci4@unibo.it

**Abstract.** I distinguish two kinds of observational advantages: (i) a given representation is observationally advantageous over another if a logical consequence of the information represented in it is observable in the former but only inferable from the latter; (ii) a given representation is observationally advantageous over another if a logical equivalence is observable in the former but only inferable from the latter. The paper also discusses the following question: observing (vs inferring) a piece of information in a given representation is an advantage if the purpose of the system of representation is to directly observe what could otherwise be inferred. But if the purpose were to infer what could be otherwise be observed, then one should conversely speak of observational disadvantages.

**Keywords:** Observational advantage · Logical equivalence
Logical consequence · Diagrams · Linearity · Type · Token

This paper is a philosophical discussion of the notion of 'observational advantage' in logical representations, as it has been recently characterized [1]. Roughly, the contrast is between inferring a statement from a given representation of information, and observing that statement without inferring it. It is said that a given representation of information is observationally advantageous over another if the former allows us to observe something that can only be inferred from the latter.

For example, given the Euler diagram in Fig. 1 and the set-theoretic sentence in Fig. 2, we see that the statements '$P \cap T = Ø$' and '$R \cap T = Ø$' are both observable from the representation in Fig. 1 but not from that in Fig. 2. They are, of course, inferable from the representation in Fig. 2 by application of inference rules. Thus the representation in Fig. 1 has an observational advantage over the representation in Fig. 2. This is an informal description of the notion of observational advantage; but it will be sufficient for the purpose of the present paper.

Both '$P \cap T = Ø$' and '$R \cap T = Ø$' are logically entailed by the representation in Fig. 2; they are both 'logical consequences' of the information represented in Fig. 2. There are, however, other statements which are both observable in Fig. 1 and inferable from Fig. 2. For example, '$Q \cap P = Ø$' is logically equivalent to '$P \cap Q = Ø$'. But while both '$P \cap Q = Ø$' and '$Q \cap P = Ø$' are observable in Fig. 1, the former statement is observable in Fig. 2, but the latter is not. The latter is, of course, inferable from the information represented in Fig. 2 by application of equivalence rules (in this

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 330–335, 2018.
https://doi.org/10.1007/978-3-319-91376-6_30

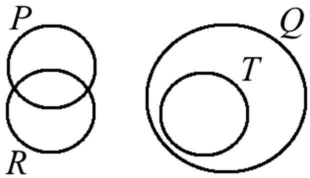


**Fig. 1.** An Euler diagram

$$\{(P \cap Q) = \emptyset, (R \cap Q) = \emptyset, T \subseteq Q\}$$

**Fig. 2.** A set-theoretic sentence

case, the rule associated with the commutative property of '∩'). Unlike 'P ∩ T = Ø' and 'R ∩ T = Ø', which are 'logical consequences' of the information represented in Fig. 2, 'Q ∩ P = Ø' is not merely a logical consequence of the information represented in it, but is more precisely a statement logically equivalent to 'P ∩ Q = Ø'. Thus, in Fig. 2 a statement logically equivalent to a statement observable in it is inferable, while in Fig. 1 the same statement is observable.

Since a statement logically equivalent to another is a logical consequence of it, while conversely not all logical consequences of a statement are logically equivalent to it, we may adopt the following terminology: a logical consequence of a statement which is *not* logically equivalent to it may be called a 'consequent', while a statement which is logically equivalent to another may be called an 'equivalent'. Since no consequent is an equivalent, we may re-phrase the characterization of observational advantages in the following terms: a given representation of information is observationally advantageous over another if the former allows us to observe a consequent or an equivalent that can only be inferred from the latter. We would then have two kinds of observational advantages: 'consequential' and 'equivalential' observational advantages.

In what sense, however, is the statement 'Q ∩ P = Ø' observable in Fig. 1 but only inferable from Fig. 2? One might object that in order to observe 'Q ∩ P = Ø' in Fig. 1 I have to 'imagine' a movement of the circles labeled Q and P, so as to have the circle Q at the left of circle P. But by the same token, the objection would go, I could directly observe 'Q ∩ P = Ø' in Fig. 2 by 'imagining' the swapping of the letters Q and P around ∩. In this sense, 'Q ∩ P = Ø' would not be more observable in Fig. 1 than in Fig. 2.

This objection is based upon an insufficient consideration of the 'sameness' of formulas in a formal language. In order to appreciate in what sense 'Q ∩ P = Ø' is observable in Fig. 1 but only inferable from Fig. 2, we have to understand in what sense 'Q ∩ P = Ø' and 'P ∩ Q = Ø' are distinct formulas which are logically equivalent (they are ‘equivalents’ in the above terminology). I distinguish between *token* and *type* formulas as follows.[1] Consider the following sentences:[2]

(1a) $P \wedge Q$
(1b) $\mathbf{P} \wedge \mathbf{Q}$

(2a) $P \wedge Q$
(2b) $Q \wedge P$

The difference between (1a) and (1b) is a typographical difference, i.e., a difference in font and size; such typographical difference is irrelevant at the syntactic level, because things such as font and size are not syntactically relevant features of the notation. Therefore, we say that (1a) and (1b) are sentence *tokens* of the same sentence *type*. In contrast, the difference between (2a) and (2b) is a difference in ordering, and since order is syntactically relevant, (2a) and (2b) are not two sentence tokens of the same sentence type, but two sentences types which are logically equivalents.

Whether couples of formulas should count as distinct tokens of the same type or as distinct types depends on what the syntactically relevant features of a notation are. This is so because the ‘sameness’ of formula types only depend on syntactically relevant features, and is unaffected by variation in syntactically irrelevant features. For example, in Euler circles the size, shape, position and orientation of the curves are not syntactically relevant features of the notation, but only the partial or total overlapping and non-overlapping of the curves are syntactically relevant. Therefore, the variation in the position of the curves by which we may obtain the Euler formula in Fig. 4 from that in Fig. 3 does not affect the ‘sameness’ of the formula type. The formulas in Figs. 3 and 4 are not distinct formula types, but only distinct formula tokens of the same type. For this reason, the formula in Fig. 3, which expresses the proposition that ‘No X is Y’ (or in set-theoretical terms, ‘$X \cap Y = \emptyset$’), also expresses the proposition that ‘No Y is X’ (or in set-theoretical terms, ‘$Y \cap X = \emptyset$’), and the same is true of the formula in Fig. 4.

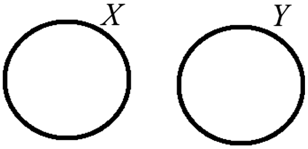


**Fig. 3.** Token of the Euler diagram type for (3a–b)

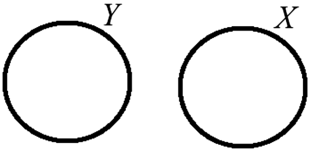


**Fig. 4.** Token of the Euler diagram type for (3a–b)

---

[1] The type/token distinction was introduced by Peirce in his *Syllabus of Logic* for the Lowell Lectures delivered in 1903 ([2] §§2.255–272), and it has since then become canonic; cf. [3].

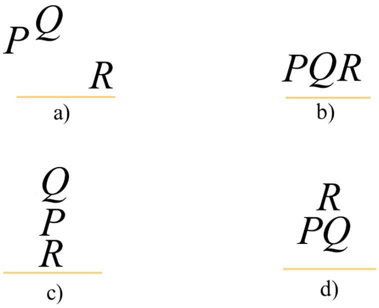
[2] Example adapted from [4].

On the contrary, in any linear language the ordering of the elements of a string is generally a syntactically relevant fact, and thus any permutation of elements in a string always produces distinct string types, and never distinct string tokens of the same string type. Thus, the two pairs of formula (3a–b) and (4a–b) contain two distinct formula types each, which are logically equivalent to one another.

(3a) No X is Y
(3b) No Y is X

(4a) $X \cap Y = \emptyset$
(4b) $Y \cap X = \emptyset$

Existential Graphs offer another example of this feature. The Alpha system of EGs is a language for the sentential calculus based on two primitives: conjunction and negation. Assertion is represented as the placement or position of the sentential variable on the sheet; conjunction is represented as the *unordered* juxtaposition of sentential variables on the sheet; negation is represented by encircling sentential variables in a closed curve (or ‘oval’). Position on the sheet and size, form and orientation of the ovals are not syntactically relevant facts, and thus variation of these features does not affect the ‘sameness’ of a graph type. Thus, each of (a)–(d) in Fig. 5 is a graph token of one and the same graph type, while (5a–f) are distinct formula types which happen to be logically equivalent:



a) b) c) d)

**Fig. 5.** Tokens of the Alpha graph type for (5a–f)

(5a) P & Q & R
(5b) P & R & Q
(5c) Q & P & R
(5d) Q & R & P
(5e) R & Q & P
(5f) R & P & Q

It is important not to conflate logical equivalence with syntactical equivalence. (4a) and (4b) are syntactically distinct formulas which happen to be logically equivalent. On the other hand, Figs. 3 and 4 are not only logically equivalent, but also syntactically equivalent (they are distinct tokens of the same type).

One might object that even in a language like that of (3a–b), (4a–b), and (5a–f), there is still the possibility of stipulating that while permutation of the elements of a string *generally* yields distinct string types, yet in some cases, e.g. in the case of symmetric relations or operations (i.e., in those cases in which permutation would not alter the logical value) permutation only yields distinct string tokens of the same type. Let us represent a symmetric relation or operator by the symbol ♠ and an anti-symmetric relation or operation by the symbol ♥, and let use the Greek letters as schematic letters. Thus, the objection would go, one could stipulate that in such a language (6a) and (6b) are distinct formula tokens of the same formula type, while (7a) and (7b) are distinct formula types.

(6a) $\xi$ ♠ $\zeta$
(6b) $\zeta$ ♠ $\xi$

(7a) $\xi$ ♥ $\zeta$
(7b) $\zeta$ ♥ $\xi$

We would thus have a notation with a dishomogeneous syntax: in such a notation, one and the same syntactical operation (permutation of elements in a string) would not yield invariably the same syntactical result: with ♠ permutation would yield distinct tokens of the same type, while with ♥ it would yield distinct types. The difference would, moreover, be dependent on semantics: for the difference between the outcome of permutation of the elements flanking ♠ and the outcome of permutation of the elements flanking ♥ would be due to the difference in *meaning* between ♠ and ♥ (the former being a symmetric relation or operation, the latter being antisymmetric). In other words, such a notation would be the outcome of a *systematic conflation* between syntax and semantics, or more precisely between syntactical equivalence and semantical equivalence. In order to avoid such a conflation, it is only necessary to recognize that in any linear notation whatever, permutation always yields distinct types, and that in order to express the differences in meaning between symmetric and antisymmetric relations rules of logical equivalence are introduced which would distinguish the outcome of the permutation of the elements flanking ♠ (distinct types which are logically equivalents) from the outcome of the permutation of the elements flanking ♥ (distinct types which are not logically equivalents).

Since the possibility of syntactically dishomogeneous linear notations is thus excluded, we can characterize linear notations in general as those in which permutation always yields formula types and never formula tokens. In a linear notation, then, the information conveyed by a formula resulting by permuting the elements of another formula is not observable in the latter, but only inferable from it by means of appropriate logical rules (like the rule of commutation). Conversely, in a non-linear notation the information conveyed by a formula resulting by permuting the elements of another formula can be observable in the latter, for the permutation in this case may yield distinct tokens of the same type. Since, then, in a non-linear notation one and the same formula token corresponds to what in a linear notation would count as distinct formula types, and since as we have seen a given representation of information is said to be observationally advantageous over another if the former allows us to observe a consequent ('consequential' observational advantage) or an equivalent ('equivalential' observational advantage) that can only be inferred from the latter, it follows that in general *non-linear notations are 'equivalentially' observationally advantageous over linear ones*, for in the former it is possible to observe what in the latter can only be inferred (by application, e.g., of commutation rules).

Stapleton et al. claim that 'representations of information that allow statements to be observed as true, without the need for inference, can be considered advantageous' [1, p. 145]. We should speak of 'advantages' only in relation to a purpose. If the purpose of devising or using a notation is to facilitate the drawing of inferences, then a notation that allows statements to be observed without the need for inference is certainly advantageous (for that purpose). But if on the contrary the purpose of devising or using a notation is to analyze inference itself (as contrasted to facilitating it), then it is by no means obvious that the possibility of observing a statement to be true without the need for inference should count as an advantage.

According to Peirce, the purpose of a notation is to analyze reasoning, not to facilitate reasoning [2, §§2.532, 3.485]. To analyze a piece of reasoning means for Peirce to dissect it into as many distinct inferential steps as possible [2, §§5.147, 3.641]. In the terms of the theory of observational advantages, this means that at any given stage of a demonstrative chain, the next step must *not* be observable from the preceding one, and rather has to be inferred from it according to the rules of the system. Euler circles allow the conclusion of a syllogism to be observable as soon as the premises are observable. By contrast, Existential Graphs do not allow this. If one's purpose is to facilitate reasoning, then the possibility that a notation allows of observing a statement to be true without the need for inference is an advantage of that notation; but if one's purpose it to analyze reasoning in the Peircean sense (i.e., to dissect it into as many distinct inferential steps as possible), then the possibility of observing a statement to be true without the need for inference should count as a disadvantage of that notation. Euler circles are more advantageous than other equivalently expressive notation only if considered as an instrument for the facilitation of inference. If they are considered as an instrument for the analysis of inference, where 'analysis' has to be taken in the Peircean sense, then they are more disadvantageous than, for example, Existential Graphs.

## References

1. Stapleton, G., Jamnik, M., Shimojima, A.: What makes an effective representation of information: a formal account of observational advantages. J. Logic Lang. Inf. **26**(5), 143–177 (2017)
2. Peirce, C.S.: Collected Papers. Harvard University Press, Cambridge (1933–1958)
3. Wetzel, L.: Types and tokens. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Spring 2014 edn. (2014)
4. Howse, J., Molina, F., Shin, S.-J., Taylor, J.: On diagram tokens and types. In: Hegarty, M., Meyer, B., Narayanan, N.H. (eds.) Diagrams 2002. LNCS (LNAI), vol. 2317, pp. 146–160. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-46037-3_18

# 'Diagrams': A Hybrid Visual Information Representation and Reasoning Paradigm Towards Video Analysis

Chayanika Deka Nath[1](✉) and Shyamanta M. Hazarika[2](✉)

[1] Biomimetic and Cognitive Robotics Lab, Computer Science and Engineering,
Tezpur University, Tezpur 784028, India
`cnath.nath@gmail.com`, `cnath@tezu.ernet.in`

[2] Mechanical Engineering, Indian Institute of Technology, Guwahati 781039, India
`s.m.hazarika@iitg.ernet.in`

**Abstract.** This paper presents a comprehensive representation for video analysis combining qualitative reasoning with diagrammatic reasoning. The hybrid approach is motivated by the power of diagrams that allows explicit relational representation of entities involved. Perception of qualitative information over the underlying representation, employment of inter-diagrammatic reasoning approach and their combined relevance for temporal abstractions holds key to the analysis. Activity recognition over selected videos from J-HMDB dataset are performed and encouraging results are achieved.

**Keywords:** Qualitative spatial and temporal reasoning
Diagrammatic representation and reasoning
Inter-diagrammatic reasoning

## 1 Introduction

Humans and animals perceive spatio-temporal information about spatial entities directly through vision or basic senses. Such information many a times is incomplete and imprecise. An intelligent brain manipulate these with cognitive faculties and experiences. Therefore, an intelligent vision system needs comprehensive representation with cognitive knowledge processing abilities. Qualitative spatial and temporal reasoning (QSTR) [7] is an established area boosting qualitative information abstractions over spatial substrate for everyday reasoning. QSTR can be further enriched with explicit representation power of diagrams or mental images to deduce conclusions within unique observed relations. Diagrammatic representation and reasoning (DR) [9] supports diagram based representation of a situation with manipulation ability to perceive new information.

This paper presents a hybrid approach- for powerful and expressive relational representation of video data through diagrams and reasoning via manipulation of qualitative spatio-temporal relations among tracked video objects. QSTR onto

© Springer International Publishing AG, part of Springer Nature 2018

P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 336–345, 2018.
https://doi.org/10.1007/978-3-319-91376-6_31

DR techniques are used as cognitive elements for space-time knowledge acquisition. For proof of concept, the proposed technique is evaluated on videos of J-HMDB dataset [12] for activity recognition and inspiring results are obtained.

## 2 Related Work and Motivation

*Qualitative Spatial and Temporal Reasoning:* QSTR for spatial knowledge processing is a thriving area of research. Many formalism have been developed and established: Region Connection Calculus (RCC) by Randell, Cardinal Direction Calculus (CDC) by Frank, Rectangle Algebra (RA) by Balbiani and Allen's interval logic (IA) by Allen [1] to deal with variety of spatio-temporal circumstances. A comprehensive representation, CORE9 have been forwarded by Cohn et al. for video analysis. Ah Lian Kor proposed an improvised hybrid cardinal direction model using RCC and CDC.

*Diagrammatic Reasoning:* Cognitive knowledge processing through analogical representation in terms of mental maps, diagrammatic representations and mental images has emerged as a promising area. REDRAW [18] a qualitative structure analyzer is based on diagram manipulation with logical reasoning. Anderson [2] defined inter-diagrammatic reasoning (IDR) for spatio-temporal abstractions over defined 'diagram' based representation. Narayanan, proposed abstraction of motion based relations over spatial structure through predefined knowledge.

*Motivation:* Power of heterogeneous framework, combining diagrams and formal logic has influenced multidisciplinary research findings; mathematical theorem prover [19], spatial problem solver [3]. A combined diagrammatic and sentential representation is suggested by Gottfried in [10,11] to enrich QSTR for results within confined relational subset. Freska [8] introduced need of comparison between formal and DR processes for same underlying problems. [13] established conceptual knowledge as a common language to generalize formal and diagrammatic approaches. Motivated by these facts, the authors aim to unify conceptual and formal problem solving techniques for video data analysis. In recent work QSTR over DR techniques are employed for motion event detection and activity recognition [14,15]. This paper focus to elaborate QSTR onto DR techniques for a comprehensive representation and reasoning for video data analysis supported by activity recognition. Interval relations [1] over DR are exploited to provide a common visual background for spatial and temporal knowledge acquisition.

## 3 'Diagrams': The Proposed Methodology

The paper forwards a hybrid representation and reasoning paradigm of video data through defined diagrams. *Diagrams* are image matrix representation of video frames in a 2-D frame of reference (x-y axes coinciding camera axes) with tracked objects, their properties and relations. QSTR and DR techniques are implemented on diagrams through methodologies involving diagram creation, perceptual and diagram modification for spatio-temporal knowledge acquisition. Figure 1 shows a conceptual architecture of the proposed approach. The ‘diagram’ components and defined methodologies are illustrated in this section.

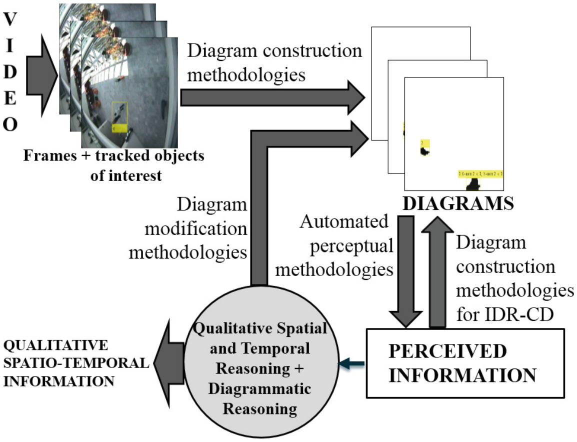
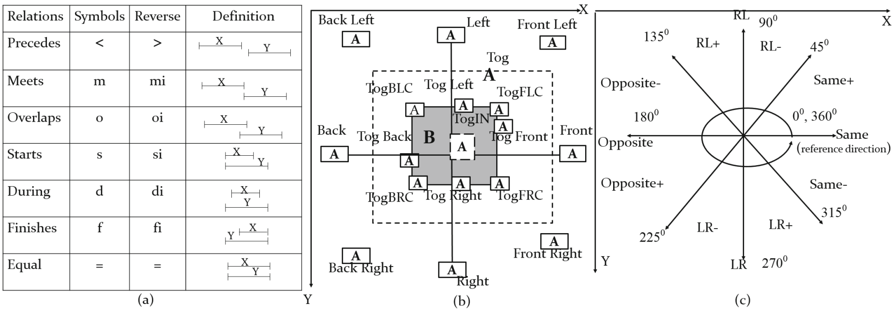


**Fig. 1.** Conceptual architecture of proposed methodology.

*Diagrammatic Objects:* Diagrammatic objects are considered to be of three types. Tracked video objects of interest are represented as ‘closed polygons’, distance among tracked object pairs as ‘lines’ and their direction of displacement during motion are represented as ‘rays’. Object’s properties such as, minimum and maximum extends of polygonal object along axes and length of line objects are maintained during their origination.

*Relations:* Along with object properties, diagrams maintain certain qualitative spatial and temporal relational information among polygon-polygon and ray-ray object pairs. These include: 13 basic Allen’s interval relations [1] along axes, 18 basic relative position relations, 3 basic relative distance relations $\{-, +, 0\}$, 8 basic displacement direction relations and relative displacement direction relations derived from $QD_8$ [4] among polygon pairs. Figure 2 shows the AI relations, relative positions and $QD_8$ relations.

*Diagram Construction Methodology:* Methodologies are defined for automatic construction of diagrams corresponding to video frames with tracked objects of interest. The procedure assigns each cell of such a diagram with specific gray intensity which are exploited by IDR-OR operator for creating *IDR-combined diagrams* (IDR-CD) for motion related information like direction of displacement. Closed polygons representing tracked objects are assigned unique pixel



| Relations | Symbols | Reverse | Definition |
|---|---|---|---|
| Precedes | < | > | X Y |
| Meets | m | mi | X Y |
| Overlaps | o | oi | X Y |
| Starts | s | si | X Y |
| During | d | di | X Y |
| Finishes | f | fi | Y X |
| Equal | = | = | X Y |

**Fig. 2.** (a) Allen's interval relations, (b) 18 basic relative position relations, and (c) 8 basic $QD_8$ relations.

intensity and the remaining pixels contain WHITE values. A sequence of diagrams called *key diagrams* ($\mathcal{K}$) are selected with difference in relative position and relative distance among interested object pairs. Each pair of $\mathcal{K}$s are sequentially combined using IDR-OR operator for IDR-CDs. IDR-CDs maintain objects and their relations as *union* of *diagrammatic objects* and *relations* in both participating key diagrams. Figure 3(a) and (b) shows a pair of diagrams automated from video frame at time point 't' and 't + 1' respectively, with IA relations and distance information as 'line' objects among polygons A, B (tracked objects). Figure 3(c) represents IDR-combined diagram of (a) and (b) with IA relations and 'rays' depicting direction of objects A and B traversing from time frame 't' to 't + 1'.

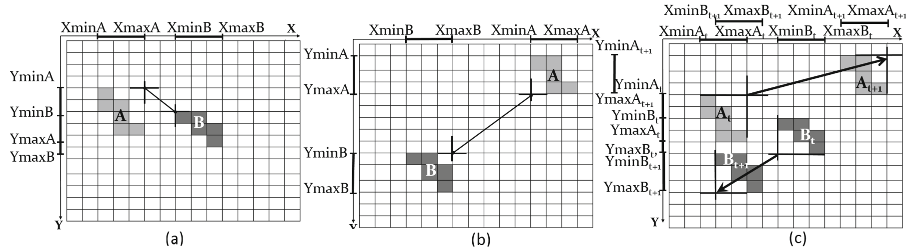
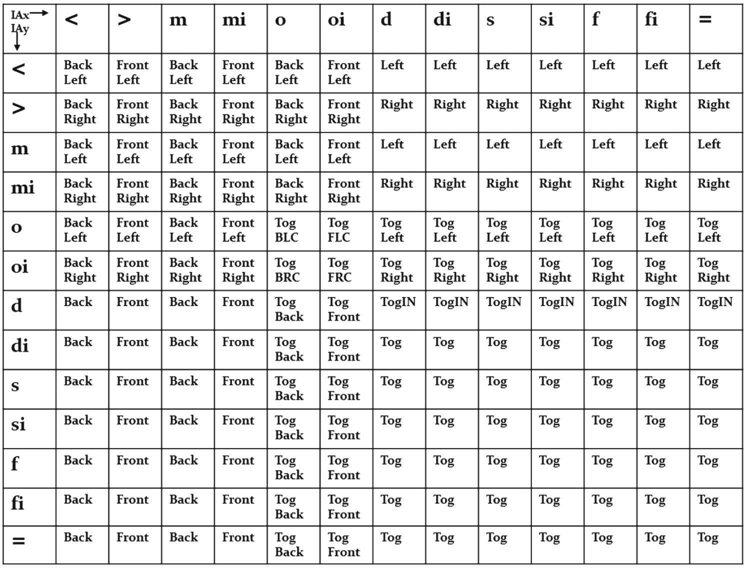


**Fig. 3.** Example video frame diagrams (a) at time point 't' (b) at time point 't + 1' with objects A, B showing distance information ('lines'), and (c) IDR-CD of (a) and (b) with displacement 'rays'.

*Automatic Perceptual Methodology:* Visual information from 'diagram' are perceived through certain perceptual mechanisms. During *diagram creation* and *modification* perceptual information are manipulated for automation of object relations. Qualitative relations are perceived through analysis of quantitative object properties. IA relations among pairs of 'polygon' along x-y axes are automated based on their extent along the axes. These *IA relations* are the core of visualizing *relative positions* and initiating diagram modification for *distance* and *displacement relations*. Figure 4 represents the defined look up table for automation of relative positions based on perceived IA relations. The length of ‘line’ objects are analyzed for relative distance relations. Displacement directions are perceived by analyzing angle between ‘ray’ objects and imaginary ‘rays’ originating at considered ray origin, parallel to x-axis; and the clockwise angle among ‘ray’ object pairs are exploited for relative displacement directions among associated ‘polygons’.



| IAx→<br>IAy↓ | **<** | **>** | **m** | **mi** | **o** | **oi** | **d** | **di** | **s** | **si** | **f** | **fi** | **=** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **<** | Back Left | Front Left | Back Left | Front Left | Back Left | Front Left | Left | Left | Left | Left | Left | Left | Left |
| **>** | Back Right | Front Right | Back Right | Front Right | Back Right | Front Right | Right | Right | Right | Right | Right | Right | Right |
| **m** | Back Left | Front Left | Back Left | Front Left | Back Left | Front Left | Left | Left | Left | Left | Left | Left | Left |
| **mi** | Back Right | Front Right | Back Right | Front Right | Back Right | Front Right | Right | Right | Right | Right | Right | Right | Right |
| **o** | Back Left | Front Left | Back Left | Front Left | Tog BLC | Tog FLC | Tog Left | Tog Left | Tog Left | Tog Left | Tog Left | Tog Left | Tog Left |
| **oi** | Back Right | Front Right | Back Right | Front Right | Tog BRC | Tog FRC | Tog Right | Tog Right | Tog Right | Tog Right | Tog Right | Tog Right | Tog Right |
| **d** | Back | Front | Back | Front | Tog Back | Tog Front | TogIN | TogIN | TogIN | TogIN | TogIN | TogIN | TogIN |
| **di** | Back | Front | Back | Front | Tog Back | Tog Front | Tog | Tog | Tog | Tog | Tog | Tog | Tog |
| **s** | Back | Front | Back | Front | Tog Back | Tog Front | Tog | Tog | Tog | Tog | Tog | Tog | Tog |
| **si** | Back | Front | Back | Front | Tog Back | Tog Front | Tog | Tog | Tog | Tog | Tog | Tog | Tog |
| **f** | Back | Front | Back | Front | Tog Back | Tog Front | Tog | Tog | Tog | Tog | Tog | Tog | Tog |
| **fi** | Back | Front | Back | Front | Tog Back | Tog Front | Tog | Tog | Tog | Tog | Tog | Tog | Tog |
| **=** | Back | Front | Back | Front | Tog Back | Tog Front | Tog | Tog | Tog | Tog | Tog | Tog | Tog |

**Fig. 4.** Relative positions of an object with respect to a reference object based on their interval relations along xy-axes.

*Diagram Modification Methodology:* During diagram creation certain information is not visually available. Diagram modification techniques were introduced with abilities to endow new information through automatic insertion of new diagrammatic objects like ‘lines’ for distance information and ‘rays’ for displacement information based on IA relations among associated polygon objects. Diagram modification techniques for ‘line’ object endpoints determination are based on IA relations among ‘polygon’ pairs along the axes. In the same way, ‘ray’ objects end points relay on IA relations along x-y axes, among considered polygons at two different time frames combined together in IDR-CDs. Figures 5 and 6 represent tables based on IA relations among polygons along axes for diagram modification to endow distance ‘lines’ and displacement direction ‘rays’.

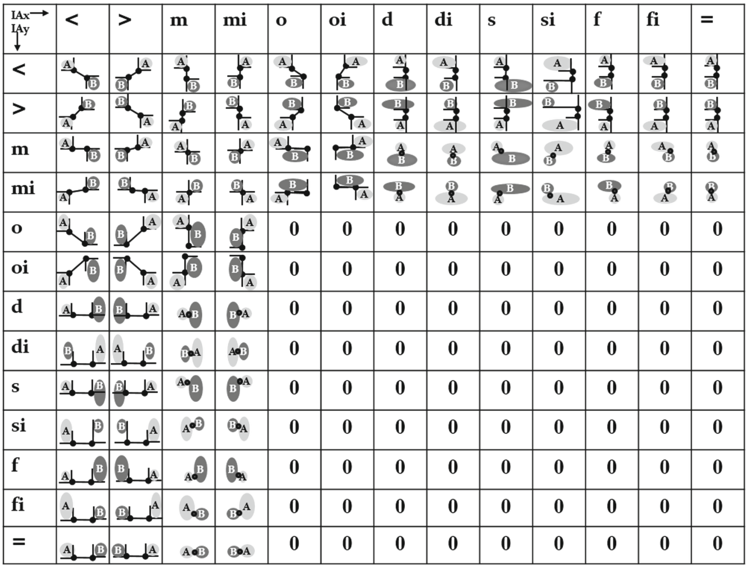


**Fig. 5.** Line object's end points determination during diagram modification for distance information between two polygons based on their interval relations along xy-axes.

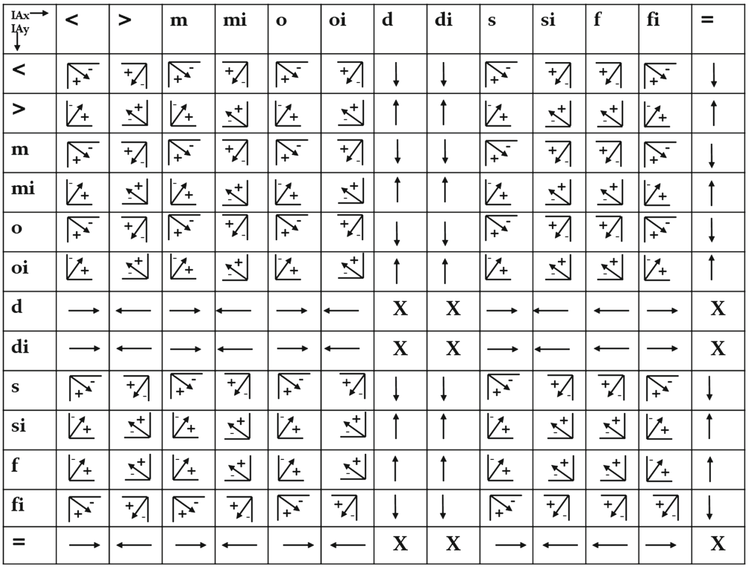


**Fig. 6.** Ray object's end points determination during combined diagram modification for displacement information of polygons from one time frame to the next based on their corresponding interval relations along xy-axes in time 't' and 't + 1'.

## 4 Application: An Example with Evaluation Details

Video analysis has tremendous applications over automatic video surveillance system, starting from event detection, extending to human activity detection, animal behavioral and movement ecology, and acquiring disaster management scenarios. This paper advocates the proposed video representation methodology through its application over video analysis for activity recognition.

The proposed technique is implemented over videos of J-HMDB dataset[1] involving selected verbs {catch, throw, shoot ball, push and pull up}. Accurate tracking is essential for a reliable video analysis. Due to various factors like, occlusion, presence of noise, various lighting conditions 100% accuracy in tracking results is itself a challenging task. For evaluation of the proposed QSTR and DR mechanism, tracking is achieved through manual labelling of objects of interest; focus is only to validate the proposed methodology, attempts at improvement of tracking being outside the scope of this work. Labelers guiding or refining labels for accuracy might end up with inaccurate tracking information which conflicts the fact about inter-dependability among accuracy in tracking and reliable video analysis. A 'diagram' sequence is automated based on tracked objects information in extracted video frames. QSTR techniques over DR are employed for automatic abstraction of spatio-temporal relations among object pairs in forward moving time in terms of displacement direction (Di) of individual objects, their relative positions (RP) at two consecutive time frames, relative displacement directions (RDi) and their relative distances (RDt). Based on these relations, sequence of certain short term activities (STA) are formulated among objects while traversing among consecutive key diagrams. This STA sequence is considered as a feature vector and a standard supervised machine (SVM) is used for associated activity classification. For example, Fig. 7 shows (a) video frames and associated key diagrams with IA relations and distance information ('lines' in blue) and (b) IDR-combined diagrams with direction of displacement information (red 'rays' depict displacement of object 1 and blue 'rays' depict displacement of object 2) in a *shoot ball* video from J-HMDB dataset. In the example, the abstracted sequence of relations among object 1 and 2 from the IDR-CDs are as shown in Table 1. Based on these relations a sequence of STAs obtained is: {AB:togIN-Tog, ABapart:TogLeftt} which constituting the minimal sequence for *shoot ball* activity among object 1 and 2. 'AB:TogIN-Tog' infers that objects A, B both are in motion and move from completely inside (TogIN) position to partial overlap (TogLeft) position; 'ABapart:TogLeft' infers that objects A, B both are in motion and move from completely partial overlap position (TogLeft) to disjoint (Left) position.

The performance of the proposed video analysis methodology for activity recognition is shown in Table 2 in terms of per class accuracy, precision, recall and F1-score. A rough comparison of recognition accuracy of the five considered activities with that of state-of-the art performances is presented. Since, performance accuracy of the activities in published state-of-the-art techniques are

---

[1] http://jhmdb.is.tue.mpg.de.

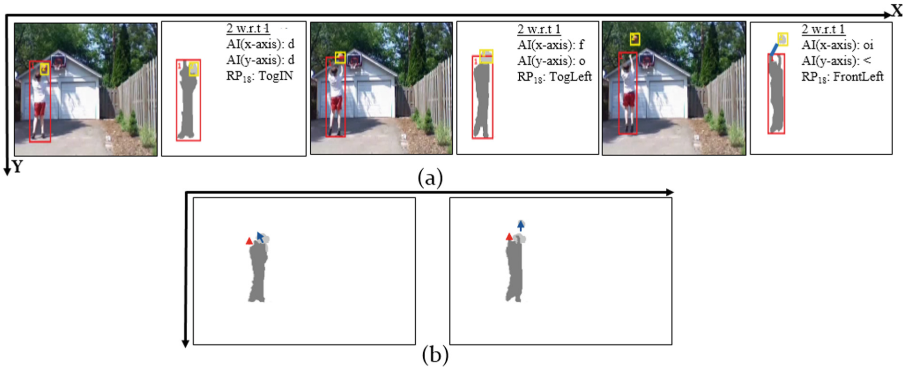


**Fig. 7.** Sequence of video frames corresponding to selected key diagram from a shoot ball video in JHMDB dataset, with (a) key diagrams with distance information and (b) IDR-CDs with displacement information. (Color figure online)

**Table 1.** Perceived qualitative relations of object A w.r.t. object B in IDR-CDs of example in Fig. 7(b).

| | Relations as a set of {Di of A, Di of B, RP of B w.r.t A in time frame 't', RP of B w.r.t A in time frame 't + 1', RDi of B w.r.t A, RDt of B w.r.t A} |
|---|---|
| **IDR-CD$_1$** | {RL, RL+, TogIN, TogLeft, Same+, 0} |
| **IDR-CD$_2$** | {RL, RL, TogLeft, Left, Same, +ve} |

**Table 2.** Activity recognition performance on videos of J-HMDB dataset reported in terms of accuracy, precision, recall and F-score. A comparison with state-of-the-art performances is presented as per class accuracy.

| | STA sequence + SVM | | | | Stacked feature vectors [16] | Pose-based CNN [6] | Dynamic programmed SVM [17] | GRP [5] | GRP w/o constraints [5] |
|---|---|---|---|---|---|---|---|---|---|
| Activities | Accuracy | Precision | Recall | F-score | Accuracy(APPROX.) | | | | |
| Catch | 87.2 | 70.0 | 70.0 | 70.0 | 38.0 | 51.0 | 22.0 | 50.0 | 56.0 |
| Throw | 95.3 | 100 | 80.0 | 88.9 | 35.0 | 38.0 | 18.0 | 72.0 | 72.0 |
| Shoot ball | 91.1 | 87.5 | 70.0 | 77.8 | 77.0 | 80.0 | 35.0 | 45.0 | 18.0 |
| Push | 97.6 | 90.9 | 100 | 95.2 | 81.0 | 95.0 | 60.0 | 75.0 | 75.0 |
| Pull up | 89.1 | 69.2 | 90.0 | 78.2 | 100 | 95.0 | 86.0 | 100 | 100 |

computed over all the 21 verbs of J-HMDB dataset a coarse comparison of the recognition performance is difficult. However, performance seems to be inspiring for the underlying video analysis methodologies. Better performance may be achieved through consideration of more relational information among STAs during recognition or via some other recognition techniques.

## 5 Conclusion

This paper focus on a comprehensive representation of video data through utilization of strength of ‘diagrams’ in human problem visualization and solution strategies. Human basically convey effective solutions based on mental maps or spatial organization of problems. The proposed methodology is a step towards formalizing the use of diagrams in cognitive vision for representation and reasoning purpose. The authors strongly advocate general concept and perception about a spatio-temporal structure to be computationally effective over formal computation with detailed and complex organizational information. A novel approach of integrating DR and QSTR techniques is being presented for video data representation in a cognitive vision system. This hybrid strategy narrows the option of ambiguity in relational composition. The work presented shows how diagrams and ‘commonsense knowledge’ can be put together for a human like problem definition through procedures like: information perception, endowing new information through diagram modification and inter diagrammatic reasoning. An application over the spatio-temporal abstractions through the proposed methodology for activity recognition is being presented. Few videos from J-HMDB dataset are being evaluated. STAs are defined over the abstracted spatio-temporal information, which serve as feature vector for a supervised machine for activity recognition. Encouraging recognition results are obtained. Further improvement could be achieved by strengthening the feature vector. As an alternative, defining a formal language automata that preserves temporal co-relations among STAs together with sequence information can be considered to uplift the framework towards precise activity recognition. This is a part of ongoing research.

## References

1. Allen, J.F.: Maintaining knowledge about temporal intervals. Commun. ACM **26**(11), 832–843 (1983)
2. Anderson, M., McCartney, R.: Diagram processing: computing with diagrams. Artif. Intell. **145**(1), 181–226 (2003)
3. Banerjee, B.: Spatial problem solving for diagrammatic reasoning. Ph.D. thesis, The Ohio State University (2007)
4. Baruah, R., Hazarika, S.: Qualitative directions in egocentric and allocentric spatial reference frames. Int. J. Comput. Inf. Syst. Ind. Manag. Appl. **6**, 344–354 (2014)
5. Cherian, A., Fernando, B., Harandi, M., Gould, S.: Generalized rank pooling for activity recognition. arXiv preprint arXiv:1704.02112 (2017)
6. Chéron, G., Laptev, I., Schmid, C.: P-CNN: pose-based CNN features for action recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3218–3226 (2015)
7. Cohn, A.G., Hazarika, S.M.: Qualitative spatial representation and reasoning: an overview. Fundam. Informaticae **46**(1), 1–29 (2001)
8. Freksa, C.: Computational problem solving in spatial substrates. Int. J. Softw. Inform. **9**(2), 279–288 (2015)

9. Glasgow, J., Narayanan, N.H., Chandrasekaran, B.: Diagrammatic Reasoning: Cognitive and Computational Perspectives. MIT Press, Cambridge (1995)
10. Gottfried, B.: Representing short-term observations of moving objects by a simple visual language. J. Vis. Lang. Comput. **19**(3), 321–342 (2008)
11. Gottfried, B.: The systematic design of visual languages applied to logical reasoning. J. Vis. Lang. Comput. **28**, 212–225 (2015)
12. Jhuang, H., Gall, J., Zuffi, S., Schmid, C., Black, M.J.: Towards understanding action recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3192–3199 (2013)
13. Lieto, A., Chella, A., Frixione, M.: Conceptual spaces for cognitive architectures: a lingua franca for different levels of representation. Biol. Inspired Cogn. Archit. **19**, 1–9 (2017)
14. Nath, C.D., Hazarika, S.M.: Combining diagrammatic reasoning with qualitative spatial and temporal reasoning for motion event detection. In: 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), pp. 1–4. IEEE (2015)
15. Nath, C.D., Hazarika, S.M.: Qualitative spatial and temporal reasoning over diagrams for activity recognition. In: Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing, p. 72. ACM (2016)
16. Peng, X., Zou, C., Qiao, Y., Peng, Q.: Action recognition with stacked fisher vectors. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 581–595. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_38
17. Soomro, K., Idrees, H., Shah, M.: Predicting the where and what of actors and actions through online action localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2648–2657 (2016)
18. Tessler, S., Iwasaki, Y., Law, K.: Qualitative structural analysis using diagrammatic reasoning. In: IJCAI (1), pp. 885–893 (1995)
19. Urbas, M., Jamnik, M.: A framework for heterogeneous reasoning in formal and informal domains. In: Dwyer, T., Purchase, H., Delaney, A. (eds.) Diagrams 2014. LNCS (LNAI), vol. 8578, pp. 277–292. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44043-8_28

# Diagrammatic Definitions of Causal Claims

Dean McHugh(✉)

Institute of Logic, Language and Computation (ILLC), University of Amsterdam,
Amsterdam, Netherlands
dean.mchugh@student.uva.nl

**Abstract.** We present a class of diagrams in which to reason about causation. These diagrams are based on a formal semantics called 'system semantics', in which states of systems are related according to temporal succession. Arguing from straightforward examples, we provide the truth conditions for causal claims that one may make about these diagrams.

## 1 Introduction

Diagrams offer a natural and highly expressive means of depicting causal relations. Flowcharts are the ubiquitous example, but even more concerted work to analyse causal relations specifically employs an abundance of visual aids. Lewis (1974, 564), for instance, diagrammatically depicts similarity orderings over worlds, while Spirtes et al. (2000) and Pearl (2009) represent Bayes nets as directed acyclic graphs.

In this paper we follow the diagrammatic tradition by presenting a class of diagrams in which to reason about causation. The bulk of the work consists in presenting a variety of cases in which diagrams represent causal relations. Regarding the underlying formal apparatus, we construct these diagrams from a semantics called 'system semantics'. In Sect. 2 we outline the approach of system semantics and see how it may be used to characterise causal relations. Section 3 provides an alternative, diagrammatic characterisation of these causal relations, and Sect. 4 refines the account by depicting two notions of 'sometimes' and 'partial' causation. In Sect. 5 we consider some further expressive power of diagrams in system semantics, showing how they may represent an agent's interaction with a system, and in Sect. 6 we conclude by outlining avenues for future work.

## 2 System Semantics for Causal Claims

To begin by analogy, system semantics aims to do for causal claims what Kripke semantics has achieved in the philosophical discussion of possible and necessary truth. Indeed, we modify Kripke semantics for modal logic to create diagrams called 'systems' that specify precisely how parts of possible worlds change through time. A system $\mathbb{S}$ is a pair $\langle St, R\rangle$ composed of a set $St$ of states and a relation $R$ of temporal succession between them. Each state represents a moment

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 346–354, 2018.
https://doi.org/10.1007/978-3-319-91376-6_32

type rather than token, and is formally a valuation of atomic sentences in propositional logic. Given states $s$ and $t$, the intuitive reading of "$s$ is related to $t$" is that, if $s$ is the current state, $t$ may be the next state after one step in time.

To illustrate, suppose we have two atomic sentences representing a switch being up ($S$), and a light being illuminated ($L$). Figure 1 represents the interaction between the switch and light. Circles depict states, accompanied by the sentences that are true at them, and arrows depict the succession relation. The diagram of Fig. 1 shows, for instance, that when the switch is up and light is off ($S, \neg L$) the system changes into the state where the switch is up and the light is on ($S, L$). And the top-left loop demonstrates that if the switch is down and light is off ($\neg S, \neg L$), then they remain so in the next state.

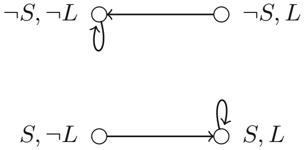


**Fig. 1.** System composed of a switch and light.

Looking at Fig. 1, it seems the following causal claim should come out true.

> The switch being up is a necessary and sufficient cause of the light being on.

The truth conditions that we propose here for such a causal claim draw on the notion of a state's *past* and *future* states. These are encoded by a system's relation of temporal succession $R$ as the states leading to and from each state along $R$. We say that the switch being up is a necessary cause of the light being on because, in every state where the light is on, we see the switch was up in the past. And the switch being up is a sufficient cause of the light being on because every state where the switch is up leads only to states where the light is on.

In general, of course, we also have to require that the above claims are not trivially satisfied, as would happen, say, if the system featured only states where the switch is up and the light is on. Triviality would result as well if the light never changed, in the sense that states where the light is off lead only to states where the light is off, and states where the light is on lead only to states where the light is on. In a slogan, then, we must additionally require that the switch being up makes some *difference* to the light being on.

We can deal with these worries of triviality by proposing a definition of minimal causation. Let us say that a state $s$ *leads to* a state $t$—and conversely, $t$ *comes from* $s$—just in case there is some path along $R$ from $s$ to $t$. Then define:

**Definition 1 (Minimal cause).** *$A$ is a minimal cause of $B$ just in case*

*(1a) some $B$-state leads to some $\neg B$-state, or vice versa, some $\neg B$-state leads to some $B$-state,*

*(1b) some A-state leads to some B state, and*
*(1c) some ¬A-state leads to some ¬B-state.*

With Definition 1 providing the minimal conditions that a causal relation must satisfy, from the point of view of system semantics, we strengthen the conditions to define the notions of necessary and sufficient causation. To do so, let us say that state $s$ *must lead to* state $t$—conversely, $t$ *must come from* $s$—just in case $t$ eventually occurs from $s$, no matter what path the system takes from $s$.

We then strengthen the notion of minimal causation like so.

**Definition 2 (Necessary cause).** *A is a necessary cause of B just in case*

*(2a) A is a minimal cause of B, and*
*(2b) every B-state must come from some A-state.*

**Definition 3 (Sufficient cause).** *A is a sufficient cause of B just in case*

*(3a) A is a minimal cause of B, and*
*(3b) every A-state must lead to some B-state.*

Condition (2b) expresses that whenever $B$ currently holds, $A$ must have held at some point in the past, no matter what path the system took to the current state. Condition (3b) expresses that whenever $A$ currently holds, $B$ will hold at some point in the future, no matter what path from the current state the system will take. The reader is invited to check that, according to Definitions 2 and 3, in the system of Fig. 1, $S$ is indeed a necessary and sufficient cause of $L$.

It turns out that Definitions (1)–(3) above can be displayed in a purely diagrammatic way, as the next section demonstrates.

## 3 A Diagrammatic Definition

Given a system $\mathbb{S}$, we diagrammatically represent condition (1a) by saying that the system in question must feature some path depicted in Fig. 2. For example, a system $\mathbb{S}$ features the topmost arrow from Fig. 2 just in case some state of $\mathbb{S}$ where both $A$ and $B$ are false leads to a state where $A$ is false and $B$ true. (For convenience we suppress '¬$A$' and '¬$B$' in Figs. 2 and 3.)

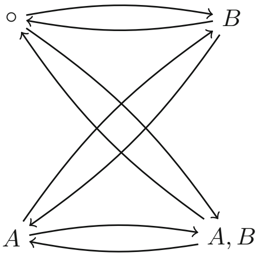


**Fig. 2.** Some $B$-state leads to some ¬$B$-state, or vice versa.

We likewise represent conditions (1b) and (1c) by means of the unshaded diagrams appearing in Fig. 3. That is, some $A$-state leads to some $B$-state in a system $\mathbb{S}$—i.e. (1b) holds—just in case $\mathbb{S}$ features some path from the bottom-right diagram of Fig. 3. And some $\neg A$-state leads to some $\neg B$-state—i.e. (1c) holds—just in case $\mathbb{S}$ features some path from the top-left diagram of Fig. 3.

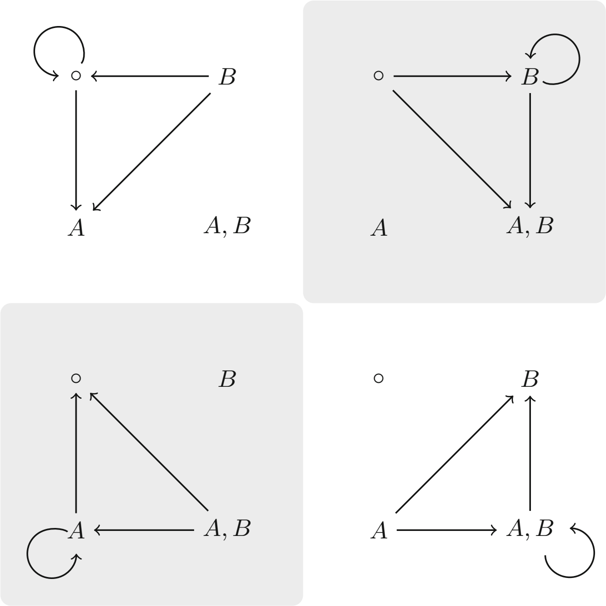


**Fig. 3.** Diagrams depicting paths from states to states.

The shaded diagrams of Fig. 3 correspond to the definitions of necessary and sufficient causation, where this time we read its arrows in terms of the 'must' mode of coming and leading. That is, a system $\mathbb{S}$ satisfies condition (2b) just in case $\mathbb{S}$ features no path from the bottom-left diagram, while condition (3b) holds in a system $\mathbb{S}$ just in case $\mathbb{S}$ features no path from the top-right diagram.

The definition of minimal causation given above is too weak on its own to serve as a definition of any intuitive notion of cause. For, conditions (1a)–(1c) only demand *some* paths of some specified kind, and so are even satisfied in systems in which states succeed one another in a completely random fashion; that is, in which every state leads to all states. In contrast, the definitions of necessary and sufficient causation are each more stringent by demanding that some paths are excluded from the system. But one might also wonder whether they are too strong to adequately capture our causal talk; there seem to be many shades of causation falling short of conditions (2b) and (3b) that our diagrams should hope to represent. In the next section we consider two less demanding ways that a causal relation—such as minimal, necessary and sufficient causation—may hold in a system. These weaker modes of causation we call ‘sometimes’ and ‘partial’ causation.

## 4 Sometimes and Partial Causation

The purpose of introducing a ‘sometimes’ modifier into causal relations is to capture causal reasoning in non-deterministic systems. Now, many analyses of causation assume that the phenomena they wish to model behave deterministically. We will not pursue the matter here, but only point out that two of the most popular analyses of causation assume some form of determinism. Firstly, Lewis’s counterfactual analysis presumes a notion of determinism in order to account for the asymmetry of causal dependence (see Menzies 2017, §2.2). Secondly, as Cartwright (1999) notes, the Bayes nets approach of Pearl (2009) and Spirtes et al. (2000) assumes determinism in order to satisfy one of their key assumptions, known as the Causal Markov Condition.

There are, nonetheless, many everyday processes we wish to model in which causes do not uniquely determine their effects. Consider, for instance, a computer with a faulty ‘on’ button, where pushing the button only sometimes succeeds in turning the computer on. (Or, more extremely, imagine the button’s success depends on some quantum set up.) This on its own is a perfectly intelligible scenario, but analyses of causation that assume determinism can only model it by introducing extraneous variables; say, by introducing a hidden variable representing the button successfully connecting with the computer. System semantics avoid such complication by allowing states to have multiple successors. Thus, in system semantics we can straightforwardly depict this scenario by means of the diagram of Fig. 4. We assume that the act of pushing the button lasts only one moment; that is, if the button is pushed at a state, then it reverts to being unpushed in the next state.

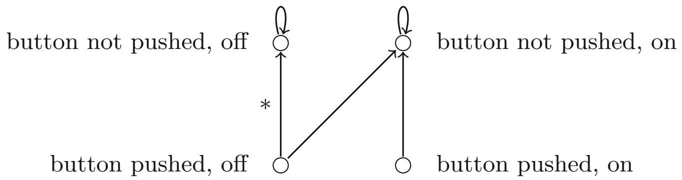


**Fig. 4.** Pushing the button sometimes causes the computer to turn on.

The non-deterministic behaviour of the button corresponds to the fact that there are two arrows coming from the state where the button is pushed and the computer is off. If the system is in that state (pushed, off), then *sometimes*—i.e. when the button did not work and the system moved along the arrow marked with a star—the computer is still off in the next state. But at other times, when the button happens to work, in the next state the computer is on.

In some cases we wish to explicitly add extra variables into our models. This occurs, say, when we want to make a background assumption explicit, or reveal the influence of a previously hidden variable. Thus, for instance, one can take into account the presence or absence of charge in the computer of Fig. 4: when there is charge ($C$), the system behaves as in Fig. 4, but when there is no charge ($\neg C$) the system always moves into a state where the computer is off. Figure 5 illustrates this new system.

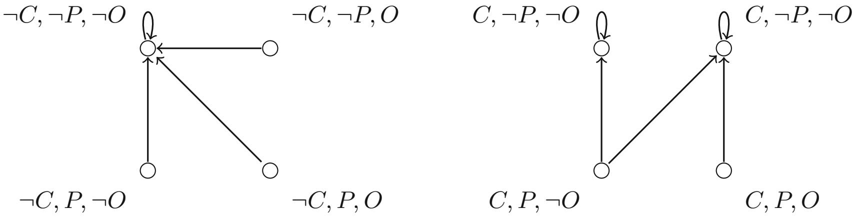


**Fig. 5.** When there is charge ($C$), pushing the button ($P$) sometimes causes the computer to turn on ($O$).

Upon examination of Figs. 4 and 5, it seems reasonable to assert the following causal claims.

(4a) In the system of Fig. 4, pushing the button is sometimes a sufficient cause of the computer turning on.

(4b) In the system of Fig. 5, when there is charge, pushing the button is sometimes a sufficient cause of the computer turning on.

In Fig. 4 we see that the path responsible for introducing the qualification 'sometimes' into (4a) is the path marked with a star. It is because of this arrow that not every path from a pushed state leads to the computer being on, meaning the system of Fig. 4 does not satisfy condition (3b). Hence, according to Definition 3, pushing the button is not a sufficient cause of the computer being on. Nonetheless, were we to restrict attention to just those times when pushing the button *is* successful—by removing the contravening arrow from the diagram—then pushing the button would be a sufficient cause of the computer turning on. This suggests the following truth condition for adding a 'sometimes' modifier to a given causal relation, defined by means of operations on diagrams.

**Definition 4 (Sometimes relation).** *A causal relation holds sometimes, in a system* $\mathbb{S}$*, just in case it holds by removing some (possibly no) arrows from* $\mathbb{S}$*.*

The system of Fig. 4 makes (4a) true since, in the system that results from removing the arrow marked with a star, pushing the button is a sufficient cause of the computer turning on.

Turning now to Fig. 5, it seems we want to say that pushing the button sometimes causes the computer to turn on, but only when there is charge. We can give the truth conditions for such conditional assertions by taking up an idea of Kratzer (1991), whereby conditionals are restrictions on quantifiers. In the present context, the proposal amounts to saying that a statement of the form 'If $A$ then $B$' is true just in case $B$ is true with respect to the $A$-states. Such a notion of conditional causal claims we call a notion of 'partial' causation, because for the causal claim to hold it need only hold in *part* of the model.

**Definition 5 (Partial relation).** *A causal relation holds partially, in a system* $\mathbb{S}$*, just in case it holds by removing some (possibly no) states from* $\mathbb{S}$*.*

Note that the definitions of sometimes and partial causation above imply that every partial relation is also a sometimes relation. For we can mimic the result of removing states from a system by removing every arrow that touches a state we wish to remove. But we cannot go the other way: there are sometimes relations that are not partial relations, as happens whenever we have to remove some but not all arrows leading from a given state. This occurs, for instance, in the system of Fig. 4 because removing any state where the computer is off—which is enough to remove the arrow marked with a star—would also make the system falsify condition (1c) and fail the test for even minimal causation.

A further advantage of depicting causal relations in terms of system semantics is that one may naturally consider multiple relations holding in the same diagram. The following section briefly outlines how such a proposal can be used to model an agent's interaction with a system.

## 5 Modelling an Agent's Interaction

By focusing on changes of states individually, rather than sentences, system semantics provides a novel level of detail absent from other approaches, notably the structural causal models of Halpern (2000) and Pearl (2009, §7.1). One advantage of the finer grain of system semantics is the abundance of ways to represent relations between states. For example, as some have demanded of automata (e.g. Baeten et al. 2011), we may naturally add succession relations to represent different kinds of change—such as those brought about by a user interacting with a system and those brought about by the system itself.

Figures 6a and b depict two different ways to add an interaction relation to the system depicted in Fig. 1. The dark lines indicate changes made by the system independently (nature's path, so to speak), while the dashed lines depict a user's path, interacting with the system. In Fig. 6a turning off the switch immediately turns off the light, whereas Fig. 6b the user takes a turn, only after which the system reacts.

Extrapolating from this simple scenario, we may model multi-agent games by introducing one relation for each agent over states of a gameplay.

(a) Interaction with immediate effect

(b) Interaction with delayed effect

**Fig. 6.** Two ways to add interaction to a system.

## 6 Conclusion

In this paper we saw some simple diagrams depict the modelling power of system semantics. Of course, one must invest quite some work just to provide a system-semantic representation of any given process, prior to analysing its causal relations. In this brief exposition we have made no argument for the capacity of the diagrams of system semantics to represent every kind of process we would wish to model. But given the widespread use of causal notions in diverse fields, such an argument would be required if system semantics for causal claims is to properly fulfil its representational ambition.

We further saw how, by encoding temporal succession into the models directly, we could analyse causal notions in a fairly straightforward manner. Of course, we have not touched upon the metaphysical issues underlying such an approach; for instance, we took the notion of temporal succession to be unproblematic. A more comprehensive appraisal of system semantics must examine whether the choices of primitives made by system semantics fare better than those of other approaches to causality, such as the assumption of a similarity ordering over worlds made by Lewis (1974). One benefit of system semantics is that its metaphysical commitment—chiefly, an ontology of states related in time—is reasonably transparent, though to fully make the case for the philosophical adequacy of system semantics, one must still argue that those are sensible commitments to make.

## References

Baeten, J.C.M., Luttik, B., van Tilburg, P.: Computations and interaction. In: Natarajan, R., Ojo, A. (eds.) ICDCIT 2011. LNCS, vol. 6536, pp. 35–54. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-19056-8_3

Cartwright, N.: Causal diversity and the Markov condition. Synthese **121**(1), 3–27 (1999)

Halpern, J.Y.: Axiomatizing causal reasoning. J. Artif. Intell. Res. **12**, 317–337 (2000)

Kratzer, A.: Conditionals. In: von Stechow, A., Wunderlich, D. (eds.) Semantics: An International Handbook of Contemporary Research, pp. 639–650. de Gruyter, Berlin (1991)

Lewis, D.: Causation. J. Philos. **70**(17), 556–567 (1974)

Menzies, P.: Counterfactual theories of causation. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Winter 2017 edn. (2017)

Pearl, J.: Causality: Models, Reasoning and Inference, 2nd edn. Cambridge University Press, New York (2009)

Spirtes, P., Glymour, C.N., Scheines, R.: Causation, Prediction, and Search. MIT Press, Cambridge (2000)

# Arcform

Benjamin Brink Allsopp(✉)

Aalborg University Copenhagen, Copenhagen, SV, Denmark
ben@it.edu

**Abstract.** Arcform is a notation for expressing diverse thoughts using nodes and arcs in a new graph-like network structure. The structure differs from directed graphs by including arcs that point from or to other arcs, and *semi* arcs where one end points from or to itself. This supports a new generative statement composition structure which allows expressive statements to be read as grammatically normal sentences while integrated into maps containing multiple statements. This paper describes this compositional structure with a special focus on a few patterns for assigning meaning to nodes and arcs that preserve the above characteristics while ensuring an even tighter integration of diverse statements into networks. A few additional features are considered before raising some far reaching questions about how it can support thought work.

**Keywords:** Network notation · Generative composition · Controlled language

## 1 Introduction

In prior work [1], arcform has been introduced as a new network notation using a new statement composition structure, requiring a new graph-like network structure. Furthermore, in that work, the compositional structure was compared to that of a range of other notations and various hypothetical ways of using triples and enclosures, the ease of using the compositional structures was tested on potential users and the expressiveness of the compositional structure was tested while capturing a text corpus in the notation.

Arcform has since been used in research projects in processes of untangling stakeholder relations and beliefs from interview data [2, 3], including one project that used the compositional structure with a language other than English [4]. It has also been considered as a foundation for an e-learning platform [5].

This paper presents the compositional structure in more detail than provided elsewhere. In the process it also presents a more specific, or regular pattern for how meaning can be captured in the compositional structure. This allows us to both explore and extend a small number of central characteristics of arcform. This version of arcform also differentiates between statement compositions (thought compositions) and object description compositions, and provides a new pattern for using the latter. Finally, a few possible implications are discussed.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 355–362, 2018.
https://doi.org/10.1007/978-3-319-91376-6_33

## 2 Objects and Thoughts

The current presentation of arcform assumes an important distinction in our thinking between objects of thought (objects) and thoughts. Objects include physical objects, imaginary physical objects, collections of objects, events, ideas, beliefs and more. In general, an object is anything that can be represented by the subject or object part of speech of normal sentences. In arcform an object is always represented by a node and labels are the simplest way of identifying an object as shown in Fig. 1. It is important to note that a thought cannot be used directly as an object.

**Fig. 1.** An object represented by a node with associated labels.

At least as important as objects are thoughts about these objects. They are also very divers and may for example include imaginings about the physical world, the experiential world or fictional words. In general, a thought in arcform is approximately anything that can be represented by a normal declarative sentence. In arcform a thought is always represented by an arc, normally with a single word label. There are many different types of thoughts in arcform, let us explore some of the most common of these.

## 3 First Order Thoughts

All first order thoughts are arcs pointing from an object and labeled with a present tense third person verb (written with brackets to accommodate two possible endings). One variation of a first order thought is an ordinary arc pointing to another object to express the same as a simple transitive sentence as shown in Fig. 2.

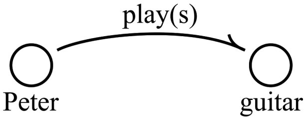


**Fig. 2.** The thought 'Peter plays guitar' represented by an arc labeled "play(s)".

The other variation of a first order thought uses a different type of arc, a semi arc that points from an object, but instead of pointing to another object, points to itself. It expresses the same as a simple intransitive sentence as shown in Fig. 3.

**Fig. 3.** The thought 'Jane sings' represented by a semi arc labeled "sing(s)".

## 4 Higher Order Thoughts

Arcform can use an arc pointing from or to another arc to represent thoughts with greater context. They are called higher order thoughts because they involve at least one other thought (a thought they point from or to). One variation of these are arcs labeled with a preposition (e.g. “for”) as shown in Fig. 4.

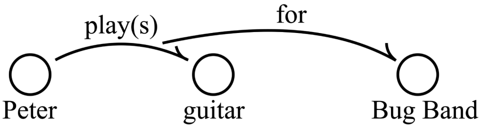


**Fig. 4.** The thought ‘Peter plays guitar for Bug Band’ represented by an arc labeled “for”.

Two other variations of higher order thoughts are arcs labelled with a subordinating conjunction (e.g. “because”) and arcs labeled with an adverb (e.g. “beautifully”). Higher order thoughts can also point from first order thoughts that are semi arcs as shown in Fig. 5.

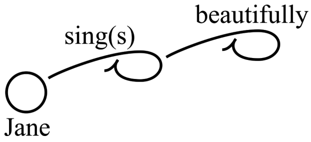


**Fig. 5.** The thought ‘Jane sings beautifully’ represented by an arc labeled “beautifully”.

Another variation of higher order thoughts is labeled with modality phrases (e.g. “do(es) not”) and point from an object to another thought as shown in Fig. 6.

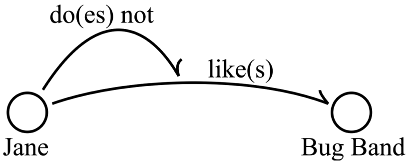


**Fig. 6.** The thought ‘Jane does not like Bug Band’ represented by the arc labeled “do(es) not”.

A central part of higher order thoughts is that they can be involved by other higher order thoughts to express new thoughts as shown in Fig. 7.

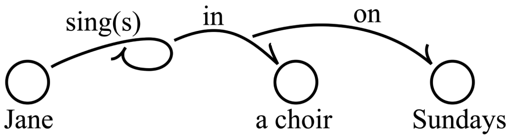


**Fig. 7.** The thought ‘Jane sings in a choir on Sundays’ involving the thought ‘Jane sings in a choir’, which involves the thought ‘Jane sings’.

This nesting of higher order thoughts within other higher order thoughts can continue indefinitely to express thoughts with greater and greater context. The possibility to do this using arcs pointing from or to other arcs differentiates arcform from all notations relying on a directed graph structure and specifically concept maps where longer expressions are created only by attaching arcs to nodes [6]. Because of this, it is worth clarifying some conventions for talking about arcform expressions:

- *Thought arcs.* Each of the displayed arcs represent a thought. For reasons that will become apparent, we can most often simply refer to these arcs as thoughts.
- *Thought compositions.* Each thought spans the tokens that it points from or to, and the tokens that these point from or to, and so on, down to include all the involved objects. We call this structure the thought composition.
- *Composition sentences.* When compositions are arranged as they are above with all thoughts pointing from left to right the labels of all the tokens can be sequenced into what we call the composition sentence.

## 5 Natural-Language-Like

Arcform is not, like a descriptive grammar, an attempt to represent how natural language syntax works (the compositional structure serves other purposes). It is also a controlled language [7] in that both the compositional structure and restrictions on how the tokens are labeled (e.g. verbs are always active present tense verbs) exclude many natural language composition sentences. However, it is meaningful to think of it as natural-language-like.

Most significantly, it has been designed to allow composition sentences to be read as grammatically normal sentences by reading the labels in sequence from left to right. Thus in Fig. 7 we put together the labels “Jane”, “sing(s)”, “in”, “a choir”, “on”, “Sundays” to read the sentence: “Jane sings in a choir on Sundays”. This exploits our familiarity with a natural language vocabulary and word ordering when interpreting thought compositions.

Furthermore, arcform seems to be expressive in the way a natural language is. Like natural languages, arcform uses a generative composition structure supporting indefinitely many compositions. Although arcform excludes many natural language sentences, different natural languages support different redundant ways of sharing the same information [8]. As long as arcform provides natural language alternatives to the excluded sentences (e.g. by using higher order thoughts to specify when something happens rather than using past tense or future tense verbs), then it maintains this expressiveness.

## 6 Arcform Maps

Above we saw that the labels of tokens could be read in sequence from left to right as normal sentences. However, this sequence is fully determined by how arcs point from and to other tokens. We can change the layout of the thought structure significantly as shown in Fig. 8 and still determine the reading order of the labels.

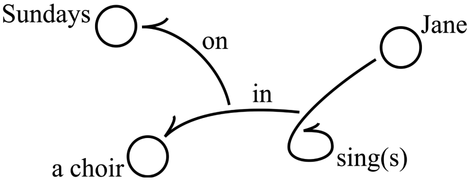


**Fig. 8.** The composition of the thought ‘Jane sings in a choir on Sundays’ with an arbitrary layout.

The obvious advantage to spreading out compositions on the plane is that multiple compositions can reuse the same tokens. This allows us to create maps integrating these compositions in networks as shown in Fig. 9. As before, every arc represents a thought, its composition is identified by the nested thoughts and objects that it involves and its composition sentence is determined by sequencing the token labels as if all arcs in the composition pointed from left to right.

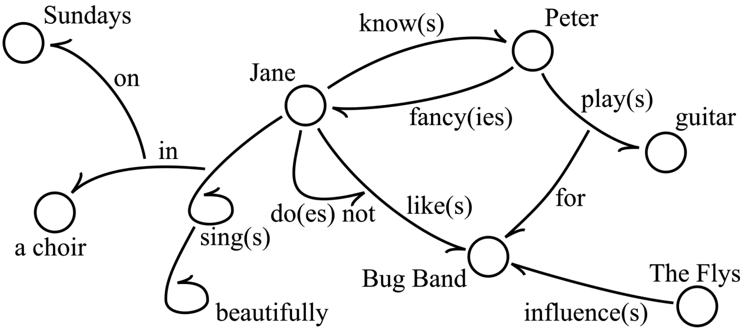


**Fig. 9.** An arcform map including multiple compositions.

## 7 Unitokenality

Prior work on arcform [1] introduced the concept of unitokenality for representations. It specifies that one meaning gets one token, no matter how many times that meaning is used in a representation. Unitokenality was seen as a requirement for a representation to be map-like. Consider how the city Edinburgh only needs to appear once in a geographical map of the world. Then contrast this to how a paragraph of natural language text may need to include multiple references to the city. Now there appear to be three increasing levels in which representations can be unitokenal with respect to our mental objects and thoughts.

1. All objects get their own token. This was the original motivation for arcform: to simply give all subjects and objects of non-trivial grammatically normal sentences their own node for reuse.
2. Some thoughts get their own token. This is seen in all prior versions of arcform. Consider how the thought ‘Jane sings’ is reused without repeating in two other thoughts in Fig. 9 above.
3. All thoughts get their own token. This requires that involved meanings in our thinking are not subsumed in other thoughts tokens.

The current version of arcform attempts to achieve the third level of unitokenality with its patterns for labels on arcs. Thought arcs do not just point from or to other meanings, but point from or to the most encompassing involved meanings. Consider how the thought ‘Jane sings in a choir’ cannot, in this version, be labeled “sings in” and point directly from the object ‘Jane’ and how this would skip giving the involved thought ‘Jane sings’ its own token. The current version of arcform ensures that ‘Jane sings’ is represented once and can be reused without repeating. This pattern is intended to apply to the composition of any involved meanings in our thinking.

## 8 Additional Features

There are also many design features of arcform that cannot be described in detail here, but strengthen the notation within the characteristics of natural-language-likeness and unitokenality, or are otherwise important to know about. An important sampling of these are: *object descriptions*, *thoughts in objects*, *conjunctions*, and *short-handing*.

Object descriptions are nodes or arcs that are neither objects or thoughts. They are a unitokenal alternatives to using labels (e.g. “a rock band” shown in Fig. 1) to describe objects. Like thoughts, descriptions also span the tokens that they points from or to and all the tokens that these point from or to and so on down. When a description composition is arranged with all its arcs pointing from left to right the labels of the tokens can be read as a grammatically normal description phrase. Figure 10 shows a description composition ‘a rock band’ with its tokens drawn with a different shade to make them easily distinguishable from objects or thoughts.

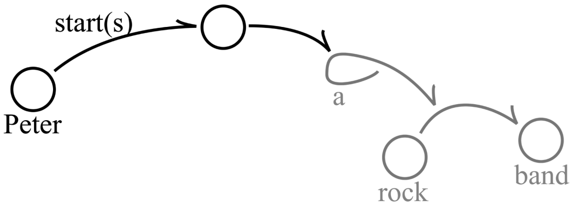


**Fig. 10.** A thought with an object linked to a description compositions.

Objects are connected to their descriptions using a *linking* thought which does not have a label, but is sometimes read as if it has the label “is” or “are”. Figure 10 shows two thoughts ‘Peter starts something’ and ‘something is a rock band’, which can be read simply as “Peter starts a rock band”. In a map the same object can have multiple descriptions and the same description can describe multiple objects.

Description composition may also involve non description tokens. For example the description ‘the king of Spain’ involves the object ‘Spain’. Likewise a description composition can also involve a thought. In Fig. 11 we can read “Peter believes that Jane likes Bug Band” where the belief is an object linked to the description ‘that Jane likes Bug Band”. The description token labelled “that” creates a description out of a thought. This pattern is important if we want to discuss claims or beliefs within arcform while allowing the involved meanings in those claims or beliefs to be unitokenally represented and reused.

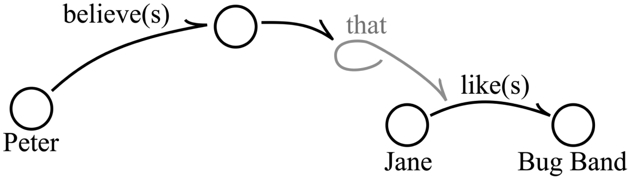


**Fig. 11.** A thought with an object described using another thought.

Finally it should be noted that arcform maps will include shapes that are not discussed here. For example there is a way of joining objects into conjunction or disjunction collections that can themselves be used as objects. There are also many ways of drawing parts of a composition in *shorthand* to hide tokens that do not at the time need to be reused. For example, in Fig. 11, we could draw a single shorthand arc labeled “believes that” from the node labeled “Peter” to the arc labeled “like(s)” when we do not need to reuse the belief ‘that Jane likes Peter’. This and many other kinds of short handing can greatly simplify many maps.

## 9 Discussion

Arcform’s design and especially its focus on expressivity, grammatically normal sentences and a high-level of unitokenality raises some interesting question about how it can support thought work. Through its seeming expressiveness we should be able to create maps on any topic, but can it go beyond simply giving a map-like experience of what would otherwise be expressed as a paragraph of text?

It seems that two maps can always be merged, here identical objects, thoughts or descriptions in the two maps become one in a new map while unique objects, thoughts or descriptions of the two maps coexist in the new map, but how far can this be taken?

Through nesting thoughts in (indefinite orders of) more contextualized thoughts, the contextualized thoughts should become less and less dependent on a containing map for their interpretation. Will these thought compositions be interpreted correctly by arcform users regardless of which other thoughts they are displayed next to? At the same time the pattern of thoughts pointing from or to the most encompassing involved meaning should allow many more opportunities for thought compositions to reuse involved thoughts. Can the combination of these support a tighter integration of more diverse thoughts then seen before? Could we in principle integrate our thoughts into one big network?

Of course, such a network of meaning would in practice require a digital implementation; at the very least to avoid the shear crowdedness of drawing it on paper. Would the composition and assignment of meaning make the expression of ideas more predictable and retrievable with structured queries? Would filtering and layout algorithms allow the dynamic and ad hoc generation of new perspectives on existing information?

Assuming the above, would the grammatically normal composition sentences make such a network more accessible for general use by non-programmers than existing knowledge base schemas? What would be the benefits of users sharing and reusing thoughts in such a network? Could it allow beliefs to be more closely connected to their counter beliefs? Could it support a new type of online social rating, not of amorphous containers of meaning like posts or pictures, but of individual thoughts?

## References

1. Allsopp, B.B.: Introducing Arc Form: Designing a satisfactory highly non-linear alternative to texts for general-purpose idea development. Ph.D. dissertation. Copenhagen: Aarhus Universitet, Institut for Uddannelse og Pædagogik (2013)
2. Tamborg, A.L., Allsopp, B.B., Fougt, S.S., Misfeldt, M.: Mapping the logics in practice oriented competence development. In: Proceedings of the 10th Congress of The European Research in Mathematics Education, Dublin, Ireland (2015)
3. Tamborg, A.L., Allsopp, B.B.: Mapping situations in implementing learning platforms. In: Brooks, A., Brooks, E., Vidakis, N. (eds.) ArtsIT 2017, DLI 2017. LNICST, vol. 229, pp. 435–444. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-76908-0_42
4. Tamborg, A.L., Allsopp, B.B.: Implementering af læringsplatforme og kulturelle logikker. Dafolo, Copenhagen (2018). (in review)
5. Allsopp, B.B.: ArcForm as a notational foundation for e-learning systems. In: Jefferies, A., Cubric, M. (eds.) Proceedings of the 14th European Conference on e-Learning. Academic Conferences and Publishing International, Reading (2015)
6. Novak, J.D., Cañas, A.J.: The Theory Underlying Concept Maps™ and How to Construct Them. Technical report IHMC CmapTools, Florida Institute for Human and Machine Cognition (2008)
7. Kuhn, T.: A survey and classification of controlled natural languages. Comput. Linguist. **40** (1), 121–170 (2014)
8. Pinker, S.: The Language Instinct. Harper Perennial Modern Classics, New York (2007)

# Euler and Venn Diagrams

# Euler Diagrams Through the Looking Glass: From Extent to Intent

Gem Stapleton[1](✉), Amirouche Moktefi[2], John Howse[1], and Jim Burton[1]

[1] Centre for Secure, Intelligent and Usable Systems, University of Brighton, Brighton, UK
`{g.e.stapleton,john.howse,j.burton}@brighton.ac.uk`
[2] Tallinn University of Technology, Tallinn, Estonia
`amirouche.moktefi@ttu.ee`

**Abstract.** Extension and intension are two ways of indicating the fundamental meaning of a concept. The extent of a concept, $C$, is the set of objects which correspond to $C$ whereas the intent of $C$ is the collection of attributes that characterise it. Thus, intension defines the set of objects corresponding to $C$ without naming them individually. Mathematicians switch comfortably between these perspectives but the majority of logical diagrams deal exclusively in extension. Euler diagrams indicate sets using curves to depict their extent in a way that intuitively matches the relations between the sets. What happens when we use spatial diagrams to depict intension? What can we infer about the intension of a concept given its extension, and vice versa? We present the first steps towards addressing these questions by defining extensional and intensional Euler diagrams and translations between the two perspectives. We show that translation in either direction leads to a loss of information, yet preserves important semantic properties. To conclude, we explain how we expect further exploration of the relationship between the two perspectives could shed light on connections between diagrams, extension, intension, and well-matchedness.

## 1 Introduction

A general term (e.g. "country", "circle", "horse", etc.) is commonly understood to refer to a collection of individuals who share one or more attributes. The set of individuals to which the term refers is called its *extension* (or *extent*), while the set of attributes shared by those individuals is called the *intension* (or *intent*) of the term. The trio (term, intent, extent) and its structure can be represented by a triangle (Fig. 1). Various names have been used in logic literature to capture this distinction. Intent is sometimes referred to as the connotation of a term while extent is said to be its denotation. Hence, a term is said to connote its intent and denote its extent. These two names can be used to characterize the two sides of our triangle that connect a term to its intent and extent. The object of this paper is to diagrammatically investigate the third side of the triangle which depicts the relation between the intent and the extent of a given term.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 365–381, 2018.
https://doi.org/10.1007/978-3-319-91376-6_34

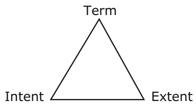


**Fig. 1.** Term-intent-extent.

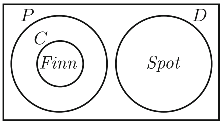


**Fig. 2.** Extension.

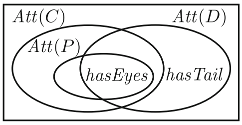


**Fig. 3.** Intension.

To illustrate, the diagram in Fig. 2 represents three sets containing invididuals: People ($P$), Children ($C$), and Dogs ($D$). It therefore presents an extensional view. The set of People, for example, comprises the individuals that have the attributes that define what it means to be a person. By contrast, Fig. 3 presents an intensional view, representing the sets of attributes of people, $Att(P)$ and so forth. We can see, from Fig. 3, that all children have all of the person attributes, including having eyes. Moreover, dogs possess attributes that no child or person has, such as having a tail. So, the dog Spot has a tail but the child Finn does not.

The distinction between intent and extent (known under various denominations) has played a significant but often undervalued role in the development of modern logic. Although it is sometimes traced in earlier writings, the distinction itself is often attributed to Antoine Arnauld and Pierre Nicole's *Logique de Port Royal* (1662). Since then, logicians who have designed logical calculi oscillated between the two interpretations. For instance, Gottfried Leibniz generally favoured the intensional interpretation while George Boole privileged the extensional interpretation, with various motivations being offered to justify the superiority of each view over the other [3]. In his 1918 survey of symbolic logic, Lewis explained the successes of Boole and his (mainly English-speaking) followers by their appeal to extent unlike their (German-speaking) predecessors who favored intent [6, pp. 35–37]. Although Gottlob Frege's logic was primarily intensional, Venn declared in 1894 that "the true intensive view is practically abandoned now, though verbally it is from time to time espoused" [13, p. 453]. Extensional logics apparently became dominant at the beginning of the twentieth century. Intensional logic was then generally believed to be at best cumbersome, if not entirely impossible [2, p. 387] [9, p. 141]. The rise and fall of intensional logics can also be traced in the development of logic diagrams. Indeed, Leibniz and several of his followers aimed at a scheme that could stand within both views, depending on whether it was the intent or the extent that was represented [1]. However, extensional diagrams shortly became dominant as geometrical relations of the diagrams appeared to match better with the logical relations of the extents than those of the intents [7]. Despite the declared superiority of extensional logic, interest in intensional logics has resurfaced in the twentieth century, notably in the footsteps of Alonzo Church and Rudolph Carnap [4]. Today, the distinction between intent and extent is commonplace and is frequently met with in modern logic textbooks [5, pp. 91–94].

There are several difficulties in determining precisely what the intent and the extent of a term involve and logicians have long considered various views on this question. Regarding intent, an objective standpoint includes any attribute possessed by the individuals to whom the term refers, whether that attribute is known or unknown to the people who use the term. A subjective view, on the other hand, supports the claim that intent should include merely the attributes that come to the mind of those who use the term. However, since different people might have different collections of attributes in mind, logic textbooks often adopt a conventional attitude in which the intent of a term refers to what is commonly attributed to it. The description of the extent of a term also involves some complications. A major difficulty concerns the definition of what counts as an individual. If we are to determine the extent of the concept "animal", should one list general species (elephants, penguins, sharks, etc.) or rather point to every specific animal considered individually? The latter technique often leads to long or infinite enumerations while the former has obvious practical advantages, although it does requires the formation of sub-classes.

The relation between intent and extent is also complex. It is first noted that, though the intent of a term might remain fixed, extent can change over time. The extent of the term "President of France" regularly changes when elections introduce new individuals with the salient attributes. Extents might also be empty; we may think, for instance, of the term "current king of France". However, equivalent extents do not necessarily indicate equivalence of intents. For example, there are various sets of attributes that can be offered to form the intent of the term whose extent contains the individuals *Spain* and *Portugal*. One might think of the term as those countries through which the *Douro River* flows. Alternatively, the term could be thought of as the countries through which the *Tagus River* flows. It could also be said that they are the last two winners to-date of the UEFA European Championship. All these intensional definitions denote the same extent, $\{Spain, Portugal\}$.

The relation between extent and intent is usually addressed through the principle of their inverse variation, whereby increasing the intent of a term by adding an attribute to it generally entails a decrease of its extent. If one thinks of the intent of the term "triangle" and adds to it the attribute of being isosceles, we remove from the extent of this term all the triangles which do not have the latter attribute (i.e. are not isosceles). Hence, the increase of intent produced a decrease of extent. In this example, the intent of a triangle is included in that of an isosceles triangle. Yet, it is the inverse that is observed for the extents, since the extent of isosceles triangles is included within the extent of triangles. It might be that an increase of intent does not produce a decrease of extent. This is the case for instance if one adds the attribute of being crossed by the *Douro River* to the intent whose attributes define the countries which are crossed by the *Tagus River*. The extent, $\{Spain, Portugal\}$, remains the same. Similarly, if a given extent is empty it does not decrease if new attributes are added to its intent. Still, in all these examples, when intent increases, extent is observed to decrease or remain stable, but definitely not increase. Similar principles are found for decreasing intents and for comparable changes in the scope of an extent.

As interesting as these principle can be, they only account for the specific case where a term is increased (respectively, decreased), meaning that it is entirely included (respectively, includes) another term. In the following, we attempt a more systematic consideration of the relations between intent and extent. In the next section we present two systems of Euler diagrams with $\otimes$-sequences: one in which curves denote the extent of terms (as is usually the case), and one in which curves denote intent. In Sect. 3 we establish a connection between interpretations of extensional and intensional diagrams, linking the semantics of the two systems through a relation that embodies the notion of one interpretation being *respectful* of another. In Sects. 4 and 5 we define translations from extensional and intensional diagrams to their respective counterparts. We show that these translations necessarily involve the loss of information yet preserve important semantic properties. Finally, we conclude with some thoughts on the implications for this work.

## 2 Syntax and Semantics

We follow a standard approach to formalizing the syntax and semantics of Euler diagrams containing $\otimes$-sequences (see Stapleton [11] for an overview of formalization techniques). We illustrate the key ideas via the example in Fig. 4. This diagram contains three *curves*, each of which has a *label*. The curves give rise to *zones*: a zone is a region of the plane inside some (possibly no) curves and outside the remaining curves. There are four zones inside this diagram, such as the one inside just the curve $P$ but outside $Q$ and $R$, and another zone outside all three curves. Sometimes, zones are *shaded*. In this example, the zone inside $R$ but outside $P$ and $Q$ is shaded. There are two $\otimes$-sequences. One of them is inside a single zone and the other comprises two $\otimes$ symbols joined by a line.

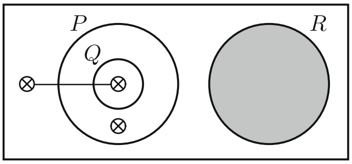


**Fig. 4.** An Euler diagram with $\otimes$-sequences.

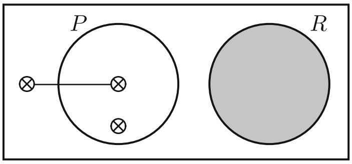


**Fig. 5.** Removing a curve: zonal regions.

We formalize diagrams at the abstract syntax level. Curves can be identified by their labels which are chosen from some given set $\mathcal{L}$. Zones are then formally defined as a pair of disjoint sets of labels, $(L_i, L_o)$. In Fig. 4, the four zones are $(\emptyset, \{P, Q, R\})$, $(\{P\}, \{Q, R\})$, $(\{P, Q\}, \{R\})$ and, lastly, the shaded zone $(\{R\}, \{P, Q\})$. The formalization of the $\otimes$-sequences is similar: they are identified at the abstract level by the set of zones in which they are placed. In our example, the two $\otimes$-sequences are $\{(\{P\}, \{Q, R\})\}$ and $\{(\emptyset, \{P, Q, R\}), (\{P, Q\}, \{R\})\}$. We now provide a formal definition of these important concepts.

**Definition 1.** *Given $\mathcal{L}$, we define a **zone over** $\mathcal{L}$ to be a pair, $(L_i, L_o)$, where $L_i \cup L_o \subseteq \mathcal{L}$, and $L_i \cap L_o = \emptyset$. The set of all zones formed from $\mathcal{L}$ is denoted $\mathcal{Z}_{\mathcal{L}}$. A set of zones formed over $\mathcal{L}$ is called a **region formed over** $\mathcal{L}$ and the set of all regions is denoted $\mathcal{R}_{\mathcal{L}}$.*

It is notable that the above definition makes the set of labels, $\mathcal{L}$, over which zones and regions are formed explicit. Later, we will be working with two distinct *systems* of diagrams that draw their labels from distinct sets. So it is important to be clear about over which label set a diagram and its components are formed.

**Definition 2.** *An **Euler diagram with** $\otimes$**-sequences**, $d_{\mathcal{L}}$, **formed over** $\mathcal{L}$ is a tuple $d_{\mathcal{L}} = (L, Z, Z^*, S)$ where*

1. *$L$ is a finite subset of $\mathcal{L}$.*
2. *$Z$ is a set of zones such that for each zone, $(L_i, L_o)$, in $Z$, $L_i \cup L_o = L$,*
3. *$Z^*$ is a subset of $Z$ whose elements are called **shaded zones**, and*
4. *$S$ is a set of regions that identify the $\otimes$-sequences: $S \subseteq \mathbb{P}Z \backslash \{\emptyset\}$.*

We will sometimes simply say Euler diagram with $\otimes$-sequences or $\mathcal{L}$-diagram, in place of Euler diagram with $\otimes$-sequences formed over $\mathcal{L}$. Furthermore, we will similarly omit saying 'formed over $\mathcal{L}$' when referring to other syntactic items.

Our attention now turns to semantics. Referring again to Fig. 4, intuitively this diagram tells us that $Q \subseteq P$ and $P \cap R = \emptyset$ due to the spatial relationships between the curves. The shading is used in the same fashion as for Venn diagrams [12], and as seen in Shin's Venn-I and Venn-II systems [10]: shaded zones represent empty sets. So, the set $R$ is empty in our example. Again, also following Shin's use of $\otimes$-sequences, regions containing an entire $\otimes$-sequence are non-empty (see Moktefi and Pietarinen [8] for the origins and development of this notation). So, we see that $P \backslash (Q \cup R) \neq \emptyset$ and, taking the universal set to be $\mathcal{U}$, $(\mathcal{U} \backslash (P \cup Q \cup R)) \cup ((P \cap Q) \backslash R) \neq \emptyset$. Given these insights, we proceed with our formalization following a standard model-theoretic approach. In our case, labels are interpreted as sets, which we then extend to interpret zones and regions:

**Definition 3.** *An **interpretation over** $\mathcal{L}$, denoted $\mathcal{I}_{\mathcal{L}}$, is a pair, $\mathcal{I}_{\mathcal{L}} = (\mathcal{U}, \Psi)$, where $\mathcal{U}$ is a set and $\Psi$ is a function, $\Psi \colon \mathcal{L} \cup \mathcal{Z}_{\mathcal{L}} \cup \mathcal{R}_{\mathcal{L}} \to \mathbb{P}\mathcal{U}$, mapping labels, zones and regions to sets such that*

1. *for each zone, $(L_i, L_o)$,*

$$\Psi(L_i, L_o) = \Big(\bigcap_{l \in L_i} \Psi(l)\Big) \backslash \Big(\bigcup_{l \in L_o} \Psi(l)\Big), \text{ and}$$

2. *for each region, $r$, $\Psi(r) = \bigcup_{z \in r} \Psi(z)$.*

The last remaining consideration, when defining the semantics, is to provide conditions under which an interpretation 'agrees' with the intended meaning of the diagram. We have already seen that shaded zones represent empty sets and that regions containing entire $\otimes$-sequences represent non-empty sets. Additionally, the relationship between the curves (in a drawn diagram) is entirely captured by the set of zones (at the abstract level). In particular, between them, all of the (abstract) zones must represent the universal set. If these three conditions are all met then the interpretation is a *model*:

**Definition 4.** *Given an Euler diagram with $\otimes$-sequences, $d_{\mathcal{L}} = (L, Z, Z^*, S)$, and an interpretation, $\mathcal{I}_{\mathcal{L}} = (\mathcal{U}, \Psi)$, we say that $\mathcal{I}$ is a* ***model*** *for $d$ provided*

1. *between them, the zones in $d_{\mathcal{L}}$ represent the universal set: $\Psi(Z) = \mathcal{U}$,*
2. *the shaded zones in $d_{\mathcal{L}}$ represent the empty set: $\Psi(Z^*) = \emptyset$, and*
3. *each $\otimes$-sequence is placed in a region that represents a non-empty set: for all $r$ in $S$, $\Psi(r) \neq \emptyset$.*

Having now defined the syntax and semantics, we introduce several further syntactic notions that will be of use later. The first focuses on the zones that are not present, (called *missing zones*), given the labels used in a diagram:

**Definition 5.** *Let $d_{\mathcal{L}} = (L, Z, Z^*, S)$ be an Euler diagram with $\otimes$-sequences. The* ***missing zones*** *of $d_{\mathcal{L}}$ are elements of*

$$\mathcal{MZ}(d_{\mathcal{L}}) = \{(L_i, L_o) \in \mathcal{Z}_{\mathcal{L}} : L = L_i \cup L_o\} \backslash Z.$$

In our running example, Fig. 4, there are four missing zones:

$$\mathcal{MZ}(d_{\mathcal{L}}) = \{(\{Q\}, \{P, R\}), (\{P, R\}, \{Q\}), (\{Q, R\}, \{P\}), (\{P, Q, R\}, \emptyset)\}.$$

Just as zones play a central role in our understanding of Euler diagrams, so too do regions that *become* zones when curves are removed. Figure 5 shows the result of removing $Q$ from Fig. 4. The zone inside just $P$ in Fig. 5 arises from two zones in Fig. 4. Regions that become zones when curves are removed are called *zonal regions* and, just like zones, can be identified by the curves that contain the region and those which do not contain the region:

**Definition 6.** *Let $IN$, $OUT$ and $L$ be sets of labels drawn from $\mathcal{L}$. A* ***zonal region*** *given $IN$, $OUT$ and $L$ is a set of zones, denoted $\langle IN, OUT, L\rangle$, where*

$$\langle IN, OUT, L\rangle = \{(L_i, L_o) : IN \subseteq L_i \wedge OUT \subseteq L_o \wedge L = L_i \cup L_o\}.$$

For example, in Fig. 4, the zonal region inside $P$ but outside $R$ is given by

$$\langle\{P\}, \{R\}, \{P, Q, R\}\rangle = \{(\{P\}, \{Q, R\}), (\{P, Q\}, \{R\})\}.$$

Having formally defined the syntax and semantics of Euler diagrams with $\otimes$-sequences, and various related notions, we are in a position to explore extensional and intensional viewpoints using Euler diagrams.

## 3 Sets of Individuals and Their Attributes

Euler diagrams are typically used to represent sets containing individuals and, therefore, visualize an *extensional* view of the world. An *intensional* viewpoint, however, considers the attributes that characterise the individuals in sets. Consider the example in Fig. 6. The diagram $d_{\mathcal{EX}}$ represents sets of individuals who are members of four sports clubs and, so, is extensional. We can see that everyone who is in the *Triathlon* club (represented by the curve labelled $T$) is also a member of the *Swimming* club. However, nobody in the *Swimming* club is in either the *Cycling* or *Football* clubs. Each club imposes the condition on its members that they must be active participants in the relevant sport; so, members of the swimming club must be able to swim, and those in the cycling and football clubs are all cyclists and, respectively, footballers. Individuals who are members of the *Triathlon* club have the attributes of being a cyclist, a swimmer and a runner, as well as the attribute of *Triathlon* club membership. As it happens, all members of the *Cycling* club are active runners. These sets of attributes therefore correspond to the *intensional* viewpoint. Therefore, $d_{\mathcal{IN}}$ (where $Int(T)$ is the set of attributes that characterise set $T$ and so forth), represents the relationships between the attributes possessed by the individuals in the sets represented by $d_{\mathcal{EX}}$, *given the particular situation just described.*

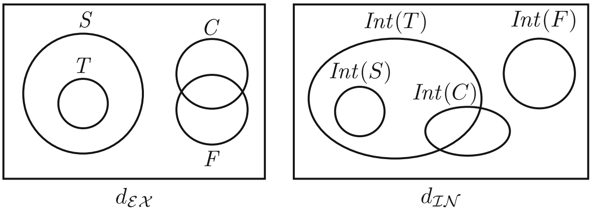


**Fig. 6.** Sports clubs and their members' attributes.

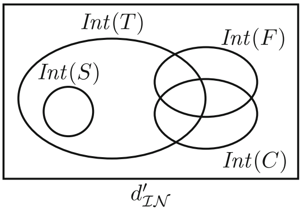


**Fig. 7.** Alternative attributes.

Whilst this example is suitable for providing intuition about the relationship between *individuals* and their *attributes*, the diagram $d_{\mathcal{IN}}$ just derived was, we emphasise, *particular* to the interpretation given. We could have had an alternative situation where all of the footballers are runners and cyclists but are not members of the triathalon or swimming clubs since none of them is able to swim. This situation would give rise to the diagram in Fig. 7. Our goal is to define a *respectful translation* from diagrams representing individuals to diagrams representing their attributes that ensures their models correspond entirely. Likewise, we also seek a respectful translation from diagrams representing attributes to diagrams representing individuals so that their models correspond. The purpose of this section is to set up a framework that allows these translations to be defined formally and for us to establish that the resulting diagrams' model sets correspond in an appropriate way.

To this end, we investigate two parallel systems of Euler diagrams with $\otimes$-sequences. We define the first system to be formed over a set of labels that we call $\mathcal{EX}$ (so $\mathcal{EX}$ is a particular choice of $\mathcal{L}$), whose elements are called **extensional labels**. Further, we define the second system to be formed over a set of labels that we call $\mathcal{IN}$, whose elements are called **intensional labels**. Importantly, we assume that $\mathcal{EX}$ and $\mathcal{IN}$ are disjoint and have the same cardinality. We now syntactically link these two systems.

**Definition 7.** *A bijective function* $Int\colon \mathcal{EX} \rightarrow \mathcal{IN}$*, which maps each extensional label to an intensional label is called an* ***intensional label allocation function****. The function* $Ext\colon \mathcal{IN} \rightarrow \mathcal{EX}$ *is the inverse of* $Int$*.*

From this point forward, we assume an intensional label allocation function, *Int*, has been defined but we also need a semantic link. $\mathcal{EX}$-diagrams and $\mathcal{IN}$-diagrams are taken to have semantics where the universal sets contain *individuals* and, respectively, *attributes*. To this end, the set of individuals is denoted $\mathcal{IND}$ and the attributes $\mathcal{ATT}$. Consequently, for example, given $\mathcal{I}_{\mathcal{EX}} = (\mathcal{U}, \Psi)$, we have $\mathcal{U} \subseteq \mathcal{IND}$. All interpretations over $\mathcal{EX}$ have universal sets that are subsets of $\mathcal{IND}$ whereas those over $\mathcal{IN}$ have universal sets that are subsets of $\mathcal{ATT}$. As our intention is to explore the relationship between information about individuals and information about their attributes, we further define a function between the sets $\mathcal{IND}$ and $\mathcal{ATT}$ to formalize this notion.

**Definition 8.** *A function,* $att\colon \mathcal{IND} \rightarrow \mathbb{P}(\mathcal{ATT})$ *is called an* ***attribute identification function****.*

As with the function *Int*, we assume from this point forward that a specific *att* is given. We now use *att* to define a link between interpretations, thus linking the semantics of the two systems:

**Definition 9.** *Let* $\mathcal{I}_{\mathcal{EX}} = (\mathcal{U}_{\mathcal{EX}}, \Psi_{\mathcal{EX}})$ *and* $\mathcal{I}_{\mathcal{IN}} = (\mathcal{U}_{\mathcal{IN}}, \Psi_{\mathcal{IN}})$ *be interpretations over* $\mathcal{EX}$ *and* $\mathcal{IN}$ *respectively. We say that* $\mathcal{I}_{\mathcal{IN}}$ *is* ***respectful*** *of* $\mathcal{I}_{\mathcal{EX}}$ *and att provided for all* $i \in \mathcal{U}_{\mathcal{EX}}$*, and for all* $P \in \mathcal{EX}$

$$i \in \Psi_{\mathcal{EX}}(P) \Leftrightarrow att(i) \supseteq \Psi_{\mathcal{IN}}(Int(P)).$$

This definition is illustrated in Fig. 8, in the case where $\Psi_{\mathcal{EX}}(P)$ is not empty (it contains $i_1$) and neither is complement (which contains $i_2$). The attributes of individual $i_1$ are $att(i_1)$ and, intuitively, since $i_1$ is in $P$ (blurring the distinction between syntax and semantics), $i_1$ must have all of the attributes that are required of individuals in $P$. In Fig. 8, this is visually indicated by the arrow from $i_1$ targeting a superset of $\Psi_{\mathcal{IN}}(Int(P))$. Likewise, the individual $i_2$ must be *missing* an attribute, say $a$, that characterises $P$. From this point forward, in general we will refer to the attributes that characterise $P$ as $P$-attributes and so forth.

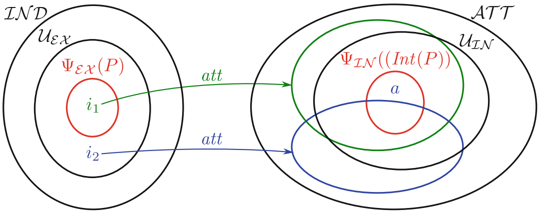
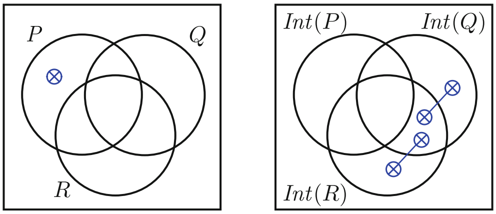


**Fig. 8.** Illustrating respectful interpretations.

## 4 From Extensional to Intensional Diagrams

Consider the example in Fig. 9, where $d_{\mathcal{EX}}$ indicates that $P \backslash (Q \cup R)$ is non-empty using an $\otimes$-sequence. Therefore, we know that there is an element in $P$ that has all of the $P$-attributes but is missing at least one of the $Q$-attributes and at least one of the $R$-attributes. So, more formally, given a model $\mathcal{I}_{\mathcal{EX}} = (\mathcal{U}_{\mathcal{EX}}, \Psi_{\mathcal{EX}})$ for $d_{\mathcal{EX}}$, $\Psi_{\mathcal{EX}}(P)$ contains an individual, say $i$, that is not in $\Psi_{\mathcal{EX}}(Q)$ nor in $\Psi_{\mathcal{EX}}(R)$.

Moreover, given a respectful interpretation, $\mathcal{I}_{\mathcal{IN}} = (\mathcal{U}_{\mathcal{IN}}, \Psi_{\mathcal{IN}})$, $att(i)$ ensures $att(i) \supseteq \Psi_{\mathcal{IN}}(Int(P))$, $att(i) \not\supseteq \Psi_{\mathcal{IN}}(Int(Q))$ and $att(i) \not\supseteq \Psi_{\mathcal{IN}}(Int(R))$. From $att(i) \not\supseteq \Psi_{\mathcal{IN}}(Int(Q))$ we can deduce that *there is a $Q$-attribute that is not a $P$-attribute*. Likewise, there is an $R$-attribute that is not a $P$-attribute, which may or may not be the same as the $Q$-attribute. This situation is captured by $d_{\mathcal{IN}}$, where two $\otimes$-sequences are placed in zonal regions. For instance, one of the $\otimes$-sequences is in $\langle \{Int(Q)\}, \{Int(P)\}, \{Int(P), Int(Q), Int(R)\} \rangle$ (i.e. the $\otimes$-sequence is inside $Int(Q)$ but outside $Int(P)$).



$d_{\mathcal{EX}}$ $d_{\mathcal{IN}}$

**Fig. 9.** A respectful translation from extensional to intensional diagrams. (Color figure online)

In general, given a single $\otimes$ placed in a zone, $(L_i, L_o)$, of an $\mathcal{EX}$-diagram we know that there is an element (in the set represented by) $(L_i, L_o)$; in what follows we frequently blur the distinction between syntax and semantics as we have just done here. This element has all of the attributes in (the sets denoted by the) intensional labels arising from $L_i$; in our previous example, the $\otimes$-sequence in $d_{\mathcal{EX}}$ was in the zone $(\{P\}, \{Q, R\})$ and, informally, had all the $P$-attributes. Importantly, such an element is missing at least one attribute from each of the intensional labels arising from $L_o$; in the previous example, informally, the $\otimes$-sequence in $d_{\mathcal{EX}}$ was missing a $Q$-attribute and an $R$-attribute. This means we know that *each intensional label* arising from $L_o$ contains an attribute that is *not* in any of the intensional labels arising from $L_i$. This leads to our next definition which identifies, for any zone, a corresponding zonal region such that if the zone in a $P$-diagram contains an $\otimes$ then the zonal region will contain an $\otimes$-sequence:

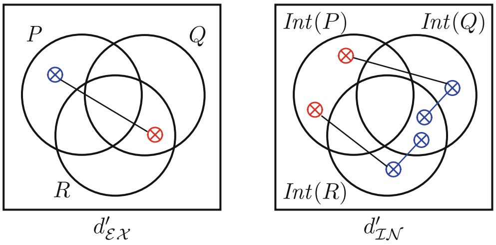
**Definition 10.** *Let $(L_i, L_o)$ be a zone formed over $\mathcal{EX}$. Let $\langle IN, OUT, I\rangle$ be a zonal region formed over $\mathcal{IN}$. Then $\langle IN, OUT, I\rangle$ is a* ***corresponding*** *$\mathcal{IN}$-****region*** *of $(L_i, L_o)$ provided*

1. *$IN$ contains a single intensional label arising from $L_o$:*
$$IN = \{Int(p)\} \text{ for some } p \in L_o,$$
2. *$OUT$ contains the intensional labels arising from $L_i$:*
$$OUT = \{Int(p) : p \in L_i\}, \text{ and}$$
3. *$I$ contains the intensional labels arising from $L_i \cup L_o$:*
$$I = \{Int(p) : p \in L_i \cup L_o\}.$$

*Given $(L_i, L_o)$, the set of zonal regions which are corresponding $\mathcal{IN}$-regions is denoted $ZR(L_i, L_o)$.*

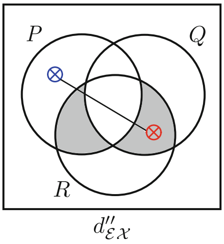
Referring again to Fig. 9, given $I = \{Int(P), Int(Q), Int(R)\}$, we have

$$ZR(\{P\}, \{Q, R\}) = \{\langle\{Int(Q)\}, \{Int(P)\}, I\rangle, \langle\{Int(R)\}, \{Int(P)\}, I\rangle\}.$$



$d'_{\mathcal{EX}}$ $d'_{\mathcal{IN}}$

**Fig. 10.** A more complex $\otimes$-sequence case. (Color figure online)



$d''_{\mathcal{EX}}$

**Fig. 11.** The impact of shading.

Whilst this gives us insight into how to translate the information provided by a single $\otimes$-sequence placed in a zone, we need to consider the more general case where $\otimes$-sequences are placed in multiple zones. Extending the example in Figs. 9 and 10, we obtain $d'_{\mathcal{IN}}$ from $d'_{\mathcal{EX}}$. Here, we have $ZR(\{P\}, \{Q, R\})$ (as given above) from the (blue) $\otimes$ in $(\{P\}, \{Q, R\})$ and

$$ZR(\{Q, R\}, \{P\}) = \{\langle\{Int(P)\}, \{Int(Q), Int(R)\}, \{Int(P), Int(Q), Int(R)\}\rangle\}$$

from the (red) $\otimes$ in $(\{Q,R\},\{P\})$. The blue $\otimes$s in $d'_{\mathcal{IN}}$ arise from the blue $\otimes$ in $d'_{\mathcal{EX}}$. Likewise for the red $\otimes$s. Given that $\otimes$-sequences provide disjunctive information, there are a range of possibilities for the presence of attributes as shown in $d'_{\mathcal{IN}}$; essentially, the $\otimes$-sequences in $d'_{\mathcal{IN}}$ capture this range of possibilities in a conjunctive normal form. Definition 11 makes this insight precise, where $r = \{z_1, \ldots, z_n\}$ can be thought of as a region containing an entire $\otimes$-sequence:

**Definition 11.** *Let $r = \{z_1, \ldots, z_n\}$ be a region formed over $\mathcal{EX}$. The elements of the set of regions, $R(r)$, given by*

$$R(r) = \{zr_1 \cup \ldots \cup zr_n : zr_1 \in ZR(z_1) \wedge \ldots \wedge \cup zr_n \in ZR(z_n)\}$$

*are **corresponding$\mathcal{IN}$-regions** of $r$.*

Having considered the presence of $\otimes$-sequences in $\mathcal{EX}$-diagrams, our attention now turns to shading. In Fig. 11, shading has been placed in the zones $(\{Q,R\},\{P\})$ and $(\{P,R\},\{Q\})$. This provides information beyond the $\otimes$-sequence, such as that $(\{Q,R\},\{P\})$ represents the empty set. Therefore, from the $\otimes$-sequence and the shading, we know that $(\{P\},\{Q,R\})$ represents a non-empty set. Thus, the information we gain about the presence of attributes arising from the $\otimes$-sequence in $d''_{\mathcal{EX}}$ reverts to what we found in Fig. 9. Consider now the shading in $(\{P,R\},\{Q\})$. This shading tells us that there are no elements in both $P$ and $R$ but outside $Q$ but does not provide any information about attributes.

Importantly, it is true in general that the *absence* of individuals in a set does not provide any information about the *absence* of attributes. This is a major point: shading in $\mathcal{EX}$-diagrams does not provide information about attributes beyond its interaction with $\otimes$-sequences. The same is true of missing zones. Having considered $\otimes$-sequences, shading, and missing zones, we are in a position to define the $\mathcal{IN}$-diagram that is a *respectful translation* of a $\mathcal{EX}$-diagram.

**Definition 12.** *Let $d_{\mathcal{EX}} = (E, Z_E, Z_E^*, S_E)$ and $d_{\mathcal{IN}} = (I, Z_I, Z_I^*, S_I)$ be Euler diagrams formed over $\mathcal{EX}$ and $\mathcal{IN}$ respectively. We say that $d_{\mathcal{IN}}$ is the **respectful translation** of $d_{\mathcal{EX}}$ given the intensional label allocation function, Int, provided:*

1. *the intensional labels in $d_{\mathcal{IN}}$ arise from the extensional labels in $d_{\mathcal{EX}}$:*

$$I = \{Int(p) : p \in E\}.$$

2. *there are no missing zones in $d_{\mathcal{IN}}$: $\mathcal{MZ}(d_{\mathcal{IN}}) = \emptyset$.*
3. *there are no shaded zones in $d_{\mathcal{IN}}$: $Z_I^* = \emptyset$.*
4. *the $\otimes$-identifiers in $d_{\mathcal{IN}}$ arise from those in $d_{\mathcal{EX}}$:*

$$S_I = \{r_{\mathcal{IN}} \subseteq Z_I : \exists r_{\mathcal{EX}} \in S_E \; r_{\mathcal{IN}} \in R_{\mathcal{IN}}(r_{\mathcal{EX}} \backslash Z_E^*)\}.$$

In Fig. 9, $d_{\mathcal{IN}}$ is the respectful translation of $d_{\mathcal{EX}}$. Likewise, $d'_{\mathcal{IN}}$ is the respectful translation of $d'_{\mathcal{EX}}$ in Fig. 10. Interestingly, $d_{\mathcal{IN}}$, Fig. 9, is the respectful translation of $d''_{\mathcal{EX}}$ in Fig. 11. This illustrates that the translation from extensional diagrams to intensional diagrams inherently loses information. Theorem 1, below, establishes that respectful translations ensure, given any model for $d_{\mathcal{EX}}$ that any interpretation which is respectful of it is a model for $d_{\mathcal{IN}}$; the theorem is illustrated in Fig. 12[1].

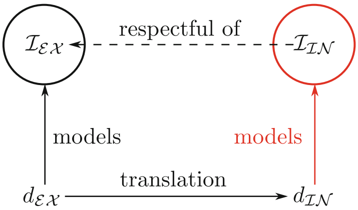


**Fig. 12.** The relationship between models.

**Theorem 1.** *Let* $d_{\mathcal{EX}} = (E, Z_E, Z^*_E, S_E)$ *be an Euler diagram formed over* $\mathcal{EX}$*. Let* $d_{\mathcal{IN}} = (I, Z_I, Z^*_I, S_I)$ *be a respectful translation of* $d_{\mathcal{EX}}$*. Let* $\mathcal{I}_{\mathcal{EX}} = (\mathcal{U}_{\mathcal{EX}}, \Psi_{\mathcal{EX}})$ *be a model for* $d_{\mathcal{EX}}$*. Let* $\mathcal{I}_{\mathcal{IN}} = (\mathcal{U}_{\mathcal{IN}}, \Psi_{\mathcal{IN}})$ *be an interpretation over* $\mathcal{IN}$ *that is respectful of* $\mathcal{I}_{\mathcal{EX}}$*. Then* $\mathcal{I}_{\mathcal{IN}}$ *is a model for* $d_{\mathcal{IN}}$*.*

## 5 From Intensional to Extensional Diagrams

Our task now is to consider what, if any, information we can derive about sets of individuals from information about attributes. We start by focusing on Fig. 13. We see that there is an attribute in $A$ that is not in $B$. However, this does not imply that there are any individuals with that attribute: the *presence* of attributes tells us nothing about the *presence* of individuals. By contrast, the *absence* of attributes does provide information about the *absence* of individuals. In our example, the shading inside $B$ but outside $A$ intuitively tells us that the attributes in $A$ include all of those in $B$. Therefore, any individual with all of the attributes in $A$ also has all of the attributes in $B$, so any such individual must also be in the set $Ext(B)$. This implies that $Ext(A)$ is a subset of $Ext(B)$, as indicated by the shading in $d_{\mathcal{EX}}$.

Having established that $\otimes$-sequences in $\mathcal{IN}$-diagrams provide no information about individuals in $\mathcal{EX}$-diagrams, our focus is now exclusively on shading and missing zones. In Fig. 14, $d'_{\mathcal{IN}}$ contains three shaded zones. From the shading in $(\{A\}, \{B, C\})$ we can see that all attributes in $A$ are all in $B$ or $C$. This implies that any individual in *both* $Ext(B)$ and $Ext(C)$ has all attributes in $A$. This insight allows us to shade the zone $(\{Ext(B), Ext(C)\}, \{Ext(A)\})$. Being in just one of $Ext(B)$ and $Ext(C)$ need not imply membership of $Ext(A)$, however.

---

[1] Proofs of Theorems 1 and 2 are omitted for reasons of space but can be found in an appendix on our website at http://readableproofs.org/looking-glass.

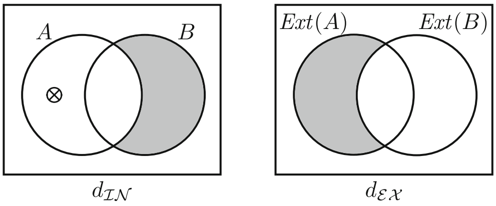


**Fig. 13.** A respectful translation from intensional to extensional diagrams.

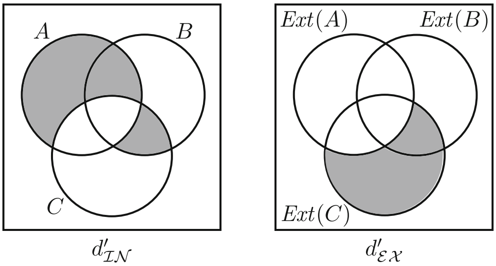


**Fig. 14.** A more complex shading case.

Consider next the shaded zone $(\{A, B\}, \{C\})$. Taking this shaded zone *in isolation* tells us that individuals with the *common* attributes of $A$ and $B$ possess all attributes in $C$. From this we cannot infer anything about the absence of individuals in the sets represented by zones of $d'_{\mathcal{EX}}$. A little more formally, from this shaded zone in $d'_{\mathcal{IN}}$, any individual, $i$, where

$$att(i) \supseteq \Psi_{\mathcal{IN}}(A) \cap \Psi_{\mathcal{IN}}(B)$$

ensures $att(i) \supseteq \Psi_{\mathcal{IN}}(C)$. But there is no zone in $d'_{\mathcal{EX}}$ whose individuals are guaranteed to have all of the attributes common to both $A$ and $B$. This is because individuals in a zone, say $(L_i, L_o)$, in $d'_{\mathcal{EX}}$, have all of the attributes in

$$\bigcup_{p \in L_i} \Psi(Int(p))$$

as opposed to an intersection of attribute sets. From this it follows that this shaded zone does not (in isolation) give rise to shading in $d'_{\mathcal{EX}}$.

However, if we consider this shaded zone *together with* the shaded zone $(\{A\}, \{B, C\})$, we form a zonal region, namely $\langle\{\{A\}, \{C\}, \{A, B, C\}\rangle$. This shaded zonal region tells us that all individuals with attributes in $A$ have all attributes in $C$: there cannot be $Ext(C)$ individuals that are not $Ext(A)$ individuals. This allows us to shade the zonal region

$$\langle\{Ext(C)\}, \{Ext(A)\}, \{Ext(A), Ext(B), Ext(C)\}\rangle$$

(of course, some of this shading in $d'_{\mathcal{EX}}$ was already obtained from the shading in $(\{A\}, \{B, C\})$). Lastly, just as taking $(\{A, B\}, \{C\})$ in isolation did not yield

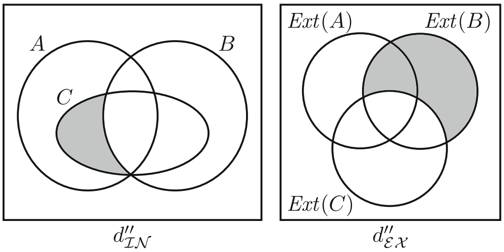


**Fig. 15.** The impact of missing zones.

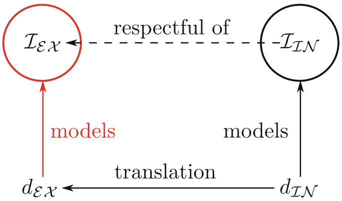


**Fig. 16.** The relationship between models.

information in $d_{\mathcal{EX}}$, the shading in $(\{B, C\}, \{A)\})$ gives rise to no information about individuals. In summary, we gain information about the absence of individuals in $\mathcal{EX}$-diagrams is when $\mathcal{IN}$-diagrams contain shading in zonal regions whose *IN* set contains a *single* label.

For our final example in the build-up to defining a respectful translation from $\mathcal{IN}$-diagrams to $\mathcal{EX}$-diagrams, we consider Fig. 15. Here, there is one shaded zone in $d''_{\mathcal{IN}}$ and also one missing zone. These two zones form a zonal region, namely $\langle\{C\}, \{B\}, \{A, B, C\}\rangle$. Consistent with our earlier examples, this gives rise to the shading in $d_{\mathcal{EX}}$. This leads to our next definition:

**Definition 13.** *Let $\langle IN, OUT, I\rangle$ be a zonal region formed over $\mathcal{IN}$ such that $|IN| = 1$. Let $\langle IN', OUT', E\rangle$ be a zonal region formed over $\mathcal{EX}$. Then $\langle IN', OUT', I\rangle$ is a **corresponding $\mathcal{EX}$-region** of $\langle IN, OUT, L\rangle$, denoted*

$$\langle IN, OUT, I\rangle \equiv_c \langle IN', OUT', E\rangle$$

*provided*

1. *$IN'$ contains a label arising from OUT: $IN' = \{Ext(a) : a \in OUT\}$,*
2. *$OUT'$ contains the label arising from IN: $OUT' = \{Ext(a) : a \in IN\}$, and*
3. *E contains the intensional labels arising from I: $E = \{Ext(a) : a \in I\}$.*

We are now in a position to define a respectful translation from $\mathcal{IN}$-diagrams to $\mathcal{EX}$-diagrams:

**Definition 14.** *Let $d_{\mathcal{IN}} = (I, Z_I, Z_I^*, S_I)$ and $d_{\mathcal{EX}} = (E, Z_E, Z_E^*, S_E)$ be Euler diagrams formed over $\mathcal{IN}$ and $\mathcal{EX}$ respectively. We say that $d_{\mathcal{EX}}$ is the **respectful translation** of $d_{\mathcal{IN}}$ given the intensional label allocation function Int provided:*

1. *the extensional labels in $d_{\mathcal{EX}}$ arise from the intensional labels in $d_{\mathcal{IN}}$: $E = \{Ext(a) : a \in I\}$.*
2. *there are no missing zones in $d_{\mathcal{EX}}$: $\mathcal{MZ}(d_{\mathcal{EX}}) = \emptyset$.*

3. *the shaded zones $d_{\mathcal{EX}}$ arise from some shaded and missing zones in $d_{\mathcal{IN}}$:*

$$\begin{gathered} Z_E^* = \{z_E \in \langle IN', OUT', E\rangle : \exists \langle IN, OUT, I\rangle \subseteq Z_I^* \cup \mathcal{MZ}(d_I) \\ |IN| = 1 \wedge \langle IN, OUT, I\rangle \equiv_c \langle IN', OUT', E\rangle\}, \end{gathered}$$

4. *there are no $\otimes$-identifiers in $d_{\mathcal{EX}}$: $S_E = \emptyset$.*

In Fig. 13, $d_{\mathcal{IN}}$ is the respectful translation of $d_{\mathcal{EX}}$, which illustrates that the translation from $\mathcal{IN}$-diagrams to $\mathcal{EX}$-diagrams inherently loses information provided by $\otimes$-sequences. We have a similar situation with shading, where $d'_{\mathcal{IN}}$ is the respectful translation of $d'_{\mathcal{EX}}$ in Fig. 14; here the shading could only be partially translated. Lastly, $d''_{\mathcal{IN}}$ in Fig. 15 highlights the role of missing zones when respectfully translating to $d_{\mathcal{EX}}$. Theorem 2, below, establishes that respectful translations ensure that models for $d_{\mathcal{IN}}$ respect only interpretations which are models for $d_{\mathcal{EX}}$; the theorem is illustrated in Fig. 16.

**Theorem 2.** *Let $d_{\mathcal{IN}} = (I, Z_I, Z_I^*, S_I)$ be an Euler diagram formed over $\mathcal{IN}$. Let $d_{\mathcal{EX}} = (E, Z_E, Z_E^*, S_E)$ be a respectful translation of $d_{\mathcal{IN}}$. Let $\mathcal{I}_{\mathcal{IN}} = (\mathcal{U}_{\mathcal{IN}}, \Psi_{\mathcal{IN}})$ be a model for $d_{\mathcal{IN}}$. Let $\mathcal{I}_{\mathcal{EX}} = (\mathcal{U}_{\mathcal{EX}}, \Psi_{\mathcal{EX}})$ be an interpretation over $\mathcal{EX}$ such that $\mathcal{I}_{\mathcal{IN}}$ is respectful of $\mathcal{I}_{\mathcal{EX}}$. Then $\mathcal{I}_{\mathcal{EX}}$ is a model for $d_{\mathcal{EX}}$.*

We now move on to summarise the results of the paper and look forward to possible directions in which it could be taken.

## 6 Conclusion

In this paper we have formalised the idea of extensional and intensional Euler diagrams, providing a systematic study of the extent and intent of a term. We established several basic results about the relationship between the two perspectives; Theorems 1 and 2 demonstrate the symmetry of the *respectfulness* relation, used to give the definitions of the translations in both directions between extent and intent. In essence our translations maintain the (minimal) information which must be true in either perspective. These results show the inevitability of information loss when translating from one system to the other.

Concerning information loss, it will be interesting to precisely characterise its nature in future work. In this context, we envisage defining an equivalence relation that syntactically characterises when two extensional diagrams (resp. intensional diagrams) give rise the same intensional diagram. Clearly, two extensional diagrams which differ *only* in their missing zones and shading give rise to the same intensional diagram (resp. extensional diagram). In Fig. 17, $d_{\mathcal{EX}1}$ and $d_{\mathcal{EX}2}$ differ in this way and both translate to $d_{\mathcal{IN}}$. In addition, $d_{\mathcal{EX}3}$, which contains an additional $\otimes$ in shaded zone, also gives rise to $d_{\mathcal{IN}}$. This is because shaded zones always represent the empty set, even if they contain an $\otimes$ symbol.

The inspiration for exploring the relationship between extensional and intensional Euler diagrams came from a thought experiment in Moktefi's 2015 paper [7]. The subject of that article is the "iconicity" of using circles to represent

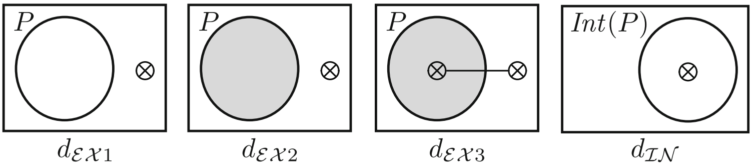


**Fig. 17.** Extensional diagrams translating to the same intensional diagram.

extent, a topic beyond scope for this discussion. However, iconicity depends on resemblances between notation and meaning; the kind of resemblance involved in this case is said to be between spatial relations of circles (e.g. one within another, two placed apart) and relations of sets (e.g. subsumption, disjointness). That is, the article [7] argues that circles are iconic to extents of terms because circles *have* the relations that sets do and that we want to depict. The thought experiment involves a hypothetical intensional Euler diagram notation without defining it, and illustrates that Euler diagrams are *less* iconic (i.e. do not possess the relevant relations) when used to depict the intent of terms.

Although this paper does not do so, the purpose of creating this formalism is to carry on the work of the thought experiment: if we accept that a notation based on circles (or, generally, closed curves) arranged in space is an effective language for reasoning about extent, what happens to this effectiveness when we use a similar notation to reason about intent? What, if anything, does the transition from extensional to intensional spatial diagrams tell us about the "effectiveness" (whether explained as iconicity, well-matchedness or using other terminology) of using space to depict extent? What form would a notation take which "has" the salient relations of intension? Would translating from extensional Euler diagrams to such an intensional notation involve the necessary loss of information? These open questions can be considered from numerous points of view, and the work we have presented is the first step towards a formal logical perspective.

## References

1. Bassler, O.B.: Leibniz on intension, extension, and the representation of syllogistic inference. Synthese **2**(116), 117–139 (1998)
2. Couturat, L.: La Logique de Leibniz. Félix Alcan, Metz (1901)
3. Dipert, R.R.: Individuals and extensional logic in schroder's 'vorlesungen uber die algebra der logik'. Mod. Log. **2–3**(1), 140–159 (1991)
4. Fitting, M.: Intensional logic (2015). https://plato.stanford.edu/archives/sum2015/entries/logic-intensional/. Accessed Dec 2017
5. Hurley, P.J.: A Concise Introduction to Logic, 12th edn. Cengage Learning, Stamford (2015)
6. Lewis, C.I.: A Survey of Symbolic Logic. University of California Press, Berkeley (1918)
7. Moktefi, A.: Is Euler's circle a symbol or an icon? Sign Syst. Stud. **43**(4), 597+ (2015)

8. Moktefi, A., Pietarinen, A.V.: On the diagrammatic representation of existential statements with Venn diagrams. J. Logic Lang. Inform. **24**(4), 361–374 (2015)
9. Shearman, A.T.: The Development of Symbolic Logic. Williams and Norgate, London (1906)
10. Shin, S.J.: The Logical Status of Diagrams. Cambridge University Press, Cambridge (1994)
11. Stapleton, G.: Delivering the potential of diagrammatic logics. In: International Workshop on Diagrams, Logic and Cognition, vol. 1132, pp. 1–8. CEUR (2013). http://ceur-ws.org/Vol-1132/paper1.pdf
12. Venn, J.: On the diagrammatic and mechanical representation of propositions and reasonings. Philos. Mag. **10**, 1–18 (1880)
13. Venn, J.: Symbolic Logic. Macmillan, Basingstoke (1894)

# Rigor and the Context-Dependence of Diagrams: The Case of Euler Diagrams

David Waszek(✉)

Institut d'Histoire et de Philosophie des Sciences et des Techniques (IHPST),
Université Paris 1 Panthéon-Sorbonne, 13 rue du Four, 75006 Paris, France
`david.waszek@gmail.com`

**Abstract.** Euler famously used diagrams to illustrate syllogisms in his *Lettres à une princesse d'Allemagne* [1]. His diagrams are usually seen as suffering from a fatal "ambiguity problem" [11]: as soon as they involve intersecting circles, which are required for the representation of existential statements, it becomes unclear what exactly may be read off from them, and as Hammer & Shin conclusively showed, any set of reading conventions can lead to erroneous conclusions. I claim that Euler diagrams can, however, be used rigorously, if they are read in conjunction with the premises they are supposed to illustrate. More precisely, I give rigorous "heterogeneous" inference rules (in the sense of Barwise and Etchemendy) – rules whose premises are a sentence and a diagram and whose conclusion is a sentence – which allow to use them safely. I conclude that one should abandon the preconception that diagrams can only be used rigorously if they can be given a context-independent semantics. Finally, I suggest that context-dependence is a widespread feature of diagrams: for instance, Mumma [12] noticed that what may be read off from a Euclidean diagram depends not only on the diagram's appearance, but also on the way it was constructed.

**Keywords:** Euler · Rigor · Context-dependence · Semantics
Heterogeneous inference

## 1 Introduction

The logical literature on diagrammatic reasoning frequently makes the implicit assumption that, in order to be used reliably, diagrams should be free-standing, i.e. have context-independent semantics and transformation rules. I shall claim that in practice, this is often not the case: diagrams are often meant to be read in conjunction with the sentences they illustrate, and this makes them neither useless nor unreliable. To defend this claim, I shall analyze a case made famous by Sun-Joo Shin's work, partly in collaboration with Hammer [6,11]: Euler diagrams. I shall show how Euler's original use of his diagrams heavily relies on contextual reading – a feature of his practice which has so far been missed by logical attempts to reconstruct it – yet is perfectly sound.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 382–389, 2018.
https://doi.org/10.1007/978-3-319-91376-6_35

## 2 Presentation of Euler's Scheme

Euler popularized the following diagrammatic scheme for representing pieces of reasoning.[1] He represents a "general notion", for instance the notion *human*, by the inside of a circle "in which one conceives all humans to be contained."[2] Relations among circles then represent relations among notions: if A corresponds to the notion *human* and B to the notion *mortal*, Fig. 1(a) displays "All humans are mortal". More generally, Euler uses this principle to illustrate the four types of proposition recognized by Aristotelian logic, namely "All A are B", "No A is B", "Some A is B" and "Some A is not B" (Fig. 1). Euler then combines these diagrams to illustrate syllogisms, i.e. inferences with two premises and one conclusion. Figure 2, for instance, illustrates the premises "All A are B" and "No B is C" and displays the conclusion "No A is C".

(a) All A are B (b) No A is B (c) Some A is B (d) Some A is not B

**Fig. 1.** Euler's diagrammatic representation of propositions [1, vol. 2, pp. 99–100]

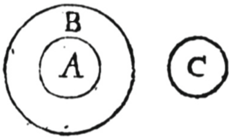


**Fig. 2.** Diagram for "All A are B; no B is C; therefore no A is C" [1, vol. 2, p. 106]

## 3 The Problem

This convention may seem clear and obvious, but there is a difficulty. As Euler remarks, the proposition "Some B is A", for instance, is compatible not just with Fig. 1(c), but also with Fig. 1(a). The issue is that Euler's diagrams always show us more than the propositions of Aristotelian logic actually claim. Figure 1(c)

---

[1] See [1, vol. 2, letters CII–CV, pp. 95–126], which is the first published version of didactic letters of Euler's from 1761 (for details, see [5] or [13, p. 417]). Similar diagrams appear before Euler in a 1661 treatise by Sturm [2, pp. 84–96] and, more systematically, in a 1712 logical text by Johann Christian Lange [3, pp. 249–268] as well as in a 1686 manuscript by Leibniz, unpublished until 1903 [4, pp. 292–321]. See e.g. [14].

[2] [1, vol. 2, p. 98]. All translations from the original French are mine.

thus shows not only “Some B is A”, but also “Some A is B”, “Some A is not B” and “Some B is not A”; similarly, Fig. 1(a) shows not only “Some B is A”, but also “Some B is not A” and “All A are B”.[3] This means that one cannot represent say “Some B is A” without deciding further questions, e.g. whether “Some A is not B”. The diagrams are *more specific* that the propositions. This problem has a well-known counterpart for geometrical diagrams: one cannot draw a triangle without drawing it with either an acute or an obtuse angle.

The same problem, namely that diagrams are more specific than propositions, resurfaces as soon as one tries to combine two premises. For instance, to illustrate “All A are B” and “Some C is A” together, Euler considers two different cases (Fig. 3) before concluding that “Some C is B”. If one only considered the first diagram, one might erroneously conclude that “All C are B”.

**Fig. 3.** Diagrams for “All A are B; some C is A; therefore some C is B” [1, vol. 2, p. 105]

As the previous example hints, a simple solution to circumvent this issue would be to represent all possible cases. One first has to consider all diagrams corresponding to each premise: for instance, for “Some A is B”, Fig. 1(c) and (a) as well as the case in which both A and B are represented by the same circle. Then, for every choice of one diagram per premise, one has to consider every possible combination of them.

But this is not what Euler does, and so, on the face of it, his method can lead to erroneous conclusions. Indeed, he never varies the representation of the premises: he always represents “Some A is B”, for instance, by Fig. 1(c). Granted: as we saw above, he does then consider the different ways in which the diagrams of the premises may be combined (Lange and Leibniz, who used similar diagrams before Euler (see footnote 1), never even do this and only consider one diagram per syllogism). But it is not enough. Take, for instance, the three diagrams Euler uses to illustrate “Some A is B” and “All C are A” (Fig. 4). Euler claims that “one cannot conclude anything, since notion C could be inside notion B entirely, or only partially, or not at all.”[4] Yet apparently, his diagrams all display “Some B is not C”! What is to stop us from drawing this false conclusion? A reasoning strategy based on the enumeration of possible cases would need to consider a diagram in which all B are A (Fig. 5). As Euler is neither drawing the erroneous conclusion nor considering this further diagram, he must be proceeding differently. But how?

---

[3] [1, vol. 2, pp. 107–108].

[4] [1, vol. 2, p. 113].

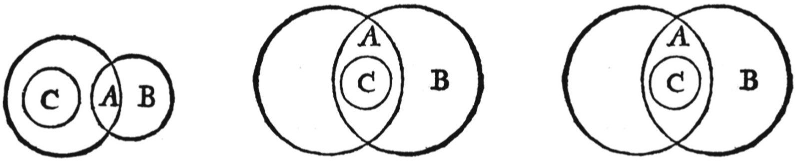


**Fig. 4.** Diagrams for “Some A is B” and “All C are A” [1, vol. 2, p. 112]

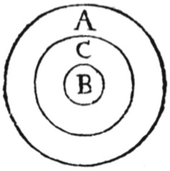


**Fig. 5.** Diagram in which “Some A is B” and “All C are A” but not “Some B is not C”

## 4 Hammer and Shin’s Diagnosis

Hammer and Shin [11] explore the possibility that Euler actually intends Fig. 1(c) (for instance) to represent “Some A is B” only, by implicitly relying on the position of letters to disambiguate his diagrams. They conclusively show, however, that no such convention can prevent all mistakes when three circles are considered together.

At any rate, it would hardly be a coherent reconstruction of Euler’s practice: Euler himself never mentions the position of letters and freely admits that his diagrams are ambiguous, without giving the slightest hint that he sees this as a problem. Clearly, for Hammer and Shin, this makes no sense: “it is quite surprising that this ambiguity did not bother Euler at all” [11, p. 5].

Hammer and Shin’s judgment relies on the assumption that diagrams can only be used rigorously if they can be given a context-independent semantics, that is, a semantics based solely on the appearance of the diagram. It is this (widespread) assumption that I would like to question here. I shall argue that Euler reads his diagrams in conjunction with the premises they are meant to illustrate. This is, I believe, quite typical of the informal use of diagrams: their appearance is not self-sufficient; to make sense of them, one has to look at them in a certain way.

## 5 Another Analysis of Euler’s Practice

To clarify Euler’s own practice with his diagrams, let us turn to letter CV, where he explains his procedure again, this time on an example of the form “No A is B; some B is C; therefore some C is not A”. For “No A is B”, he uses Fig. 1(b). For “Some B is C”, however, he adds a star to his diagram (Fig. 6), and explains the syllogism thus:

> Since a part of space C is in B and since all of space B is outside space A, it is obvious that the same part of space C must also be outside space A [...]
> One should pay careful attention to the fact that this conclusion only concerns the part * of notion C which is immersed in notion B. For the rest [of notion C], it is uncertain whether it too is excluded from notion A, as in [Fig. 7(a)], or whether it is contained in it entirely, like [Fig. 7(b)] or only in part as in [Fig. 7(c)]. Since this is uncertain, the rest of space C does not enter into consideration at all; the conclusion is restricted to what is certain, namely, that the same part of space C which is contained in space B is certainly outside space A [...].[5]

Euler is not introducing a new, altered system here; if he was, he would at least attempt to explain systematically how to use the new star for other kinds of statements. Rather, he is attempting to make explicit the way he understands his procedure. In this context, the star is a way to draw attention to the relevant part of the diagram, or more precisely, to explain how the diagram should be looked at.

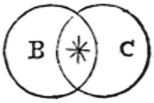
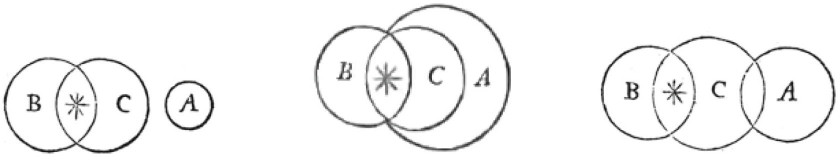


**Fig. 6.** Euler's starred version of his diagram for "Some B is C" [1, vol. 2, p. 121]



(a) (b) (c)

**Fig. 7.** Starred diagrams for "No A is B; some B is C; therefore some C is not A" [1, vol. 2, pp. 121–122]. In diagram (b), the border of region A is not a circle: regions A and B are supposed not to intersect.

As such, Euler's explanations provide us with a crucial hint: he uses his diagrams *jointly with the premises they are meant to illustrate*. When looking at his circle diagrams, he keeps in mind which regions he knows something about. For example, in the quote above, he explains that regions other than the starred one cannot support inferences. Drawing all possible diagrams is then unnecessary: one can draw just one diagram, and rely only on what is general about it. I shall argue that this procedure allows Euler to reason with his diagrams without risk of error.

---

[5] [1, vol. 2, pp. 121–122].

## 6 Making it Precise: Heterogeneous Inference Rules

To make this analysis of Euler's procedure precise, I shall formulate explicit *heterogeneous* inference rules which I take to provide a fairly faithful reconstruction of Euler's practice.

First, a clarification about the phrase "heterogeneous inference". Barwise and Etchemendy championed the idea of formal systems based on both sentences and diagrams, the first example of which was provided by their own *Hyperproof* system.[6] The point of such systems was that they permitted heterogeneous inference rules: for instance, rules whose premise was a diagram and whose conclusion was a sentence. This allowed a division of inferential labor between text and diagram, under the assumption that some inferences would be best approached diagrammatically and others sententially. What I shall suggest here is to use heterogeneous inference rules having a sentence and a diagram as premises and a sentence as conclusion, in order to model Euler's contextual use of diagrams.

Now, let us start from two syllogistic premises and a diagram with three labeled regions. The heterogeneous inference rules are the following. The first two only rely on the diagram; the last four rely both on the diagram and on the premises it illustrates.[7]

1. From the diagram, infer "All A is B" if it contains two regions labeled A and B, and if region A is contained in region B.
2. From the diagram, infer "No A is B" if it contains two regions labeled A and B, and if region A and region B are disjoint.
3. From the diagram and premise "Some A is C", infer "Some A is B" if the diagram contains regions labeled A, B and C and if the intersection of regions A and C is contained in region B.
4. From the diagram and premise "Some A is C", infer "Some A is not B" if the diagram contains regions labeled A, B and C and if the intersection of regions A and C is disjoint from region B.

---

[6] See in particular [8,9]. For another example in the same spirit, see Eric Hammer's heterogeneous system for Venn diagrams [7], based on Sun-Joo Shin's purely diagrammatic system [6].

[7] I argued that Euler's diagrams cannot be given a semantics on the sole basis of their appearance: they only make sense when read together with premises they illustrate. But from the point of view of these heterogeneous inference rules, which reconstruct Euler's practice by teasing apart the roles of diagrams and of premises, matters are different. Indeed, in the context of these rules, one could consider Euler's diagrams as representing universal statements only: in other words, one could give a diagram the same semantics as the conjunction of all statements "All A are B" and "No C is D" that it displays (in the sense that region A is included in region B and regions C and D do not meet). Such a semantics would make our heterogeneous rules sound. However, note that this semantics for isolated diagrams is a sort of artifact of the formal reconstruction and makes little sense from the point of view of Euler's practice, in which diagrams are never used independently from the text. In keeping with Euler's own remarks, it would be more natural to say that e.g. Fig. 1(c) is disambiguated by the context than to claim that it represents nothing at all.

5. From the diagram and premise "Some A is not C", infer "Some A is B" if the diagram contains regions labeled A, B and C and if the difference of region A and region C is contained in region B.
6. From the diagram and premise "Some A is not C", infer "Some A is not B" if the diagram contains regions labeled A, B and C and if the difference of region A and region C is disjoint from region B.

The main idea behind rules 3–6 is that an existential conclusion can only be reached on the basis of a subregion asserted to be nonempty by one of the premises. For this reason, these rules do not allow the erroneous inference from Fig. 4.

With these rules at hand, one can reconstruct Euler's practice fully. To derive a conclusion from two Aristotelian premises sharing a common term, he first draws one diagram per premise using his scheme (Fig. 1) and superposes them, producing as many diagrams as there are possible relations between the "extreme" terms (i.e. the two terms which do not appear in both premises). Finally, Euler draws a conclusion only if, according to the rules given above, each of these combined diagrams allows it.

## 7 The Drawbacks of Euler's Approach and the Later History of Logical Diagrams

I argued that Euler's practice is reliable, not that there is nothing subtle about it. Euler does not make explicit the rules governing the context-dependent use of his diagrams, and the fact is that his system is often perceived as confusing. Perhaps rules which are both implicit and contextual are intrinsically difficult, or delicate to convey; in any case, as Hammer and Shin noted [11], it is quite telling that when Euler attempts to explain precisely how he is reasoning with his diagrams, in letter CV, he resorts to enriching them (be it only as a temporary didactic device) with a new symbol – a star, reminiscent of Peirce's later crosses. In fact, what is gained when moving to the diagrams of Venn or Peirce is precisely that they are free-standing: they can be read independently of any sentential context. The diagrams then become fully independent of the text, as Euler's were not.

## 8 General Lessons

Diagrams, then, can have a context-dependent interpretation and yet be useful and reliable reasoning tools. This phenomenon has already been noticed by Mumma [12] in his work on Euclidean geometry. I drew a parallel between Euler diagrams and Euclidean diagrams above: in both cases, the diagram is often more specific than we would like; if we want to rely on it for reasoning, we need a way to distinguish those features of the diagram which are general, and those which are accidental and should not be taken into account. In the case of Euler diagrams, we saw that this requires referring back to the premises the diagram is meant to illustrate. Analogously, Mumma showed that *the mere visual inspection* *of a Euclidean diagram cannot reveal what is meant to be general about it*, and therefore what we can infer from it: we also need to know how the diagram was constructed. He tries to capture this additional information in a formal object which he calls, precisely, the *context* of the diagram.

I doubt that these cases are isolated. It is probable that most informal diagrams present this feature to some extent, as they were not designed to be free-standing. If we want to better account for the use of diagrams in practice, our logical models of reasoning should incorporate this phenomenon.

**Acknowledgments.** I would like to thank Jeremy Avigad, Ken Manders, Nicolas Michel, Marco Panza and David Rabouin for discussions on this topic.

## References

1. Euler, L.: Lettres à une princesse d'Allemagne sur divers sujets de physique & de philosophie. Imprimerie de l'Académie Impériale des Sciences, Saint Pétersbourg (1768)
2. Sturm, J.C.: Universalia Euclidea. Adrian Vlacq, Hagæ-Comitis [The Hague] (1661)
3. Lange, J.C.: Nucleus Logicae Weisianae. Henning Müller, Gissae-Hassorum [Giessen] (1712)
4. Leibniz, G.W.: Opuscules et fragments inédits de Leibniz. Ed. by L. Couturat. F. Alcan, Paris (1903)
5. Speiser, A.: Einleitung zu den Lettres à une princesse d'Allemagne. In: Euler, L.: Opera Omnia, vol. III/1, pp. VII–XLIII. Orell Füssli, Turici [Zürich] (1960)
6. Shin, S.J.: The Logical Status of Diagrams. Cambridge University Press, Cambridge (1994)
7. Hammer, E.M.: Reasoning with sentences and diagrams. Notre Dame J. Form. Log. **35**(1), 73–87 (1994)
8. Barwise, J., Hammer, E.M.: Diagrams and the Concept of Logical System. In: [10], pp. 49–78. Orig. publ. in Gabbay, D.M. (ed.) What Is a Logical System?, pp. 73–106. Clarendon Press, Oxford (1994)
9. Barwise, J., Etchemendy, J.: Heterogeneous Logic. In: [10], pp. 179–200. Orig. publ. in Glasgow, T.I., Narayanan, N.H., Chandrasekaran, B. (eds.) Diagrammatic Reasoning: Cognitive and Computational Perspectives, pp. 209–232. AAAI Press, Menlo Park and MIT Press, Cambridge (1995)
10. Allwein, G., Barwise, J. (eds.): Logical Reasoning with Diagrams. Oxford University Press, Oxford (1996)
11. Hammer, E.M., Shin, S.J.: Euler's visual logic. Hist. Philos. Log. **19**(1), 1–29 (1998)
12. Mumma, J.: Proofs, pictures, and Euclid. Synthese **175**(2), 255–287 (2010)
13. Calinger, R.S.: Leonhard Euler: Mathematical Genius in the Enlightenment. Princeton University Press, Princeton (2015)
14. Bennett, D.: Origins of the Venn diagram. In: Zack, M., Landry, E. (eds.) Research in History and Philosophy of Mathematics. PCSHPMSCPM, pp. 105–119. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22258-5_8

# Investigating Diagrammatic Reasoning with Deep Neural Networks

Duo Wang(✉), Mateja Jamnik, and Pietro Liò

Department of Computer Science and Technology, University of Cambridge,
Cambridge, UK
`wd263@cam.ac.uk`, `{mateja.jamnik,pietro.lio}@cl.cam.ac.uk`

**Abstract.** Diagrams in mechanised reasoning systems are typically encoded into symbolic representations that can be easily processed with rule-based expert systems. This relies on human experts to define diagram-to-symbol mapping and the set of rules to reason with the symbols. We present a new method of using Deep artificial Neural Networks (DNN) to learn continuous, vector-form representations of diagrams without any human input, and entirely from datasets of diagrammatic reasoning problems. Based on this DNN, we developed a novel reasoning system, Euler-Net, to solve syllogisms with Euler diagrams. Euler-Net takes two diagrams representing the premises in a syllogism as input, and outputs either a categorical (subset, intersection or disjoint) or diagrammatic conclusion (generating an Euler diagram representing the conclusion) to the syllogism. Euler-Net can achieve 99.5% accuracy for generating syllogism conclusions, and learns meaningful representations. We propose that our framework can be applied to other types of diagrams, especially the ones we are less sure how to formalise symbolically.

## 1 Introduction

Diagrams have been shown to be effective tools for humans to represent and reason about complex concepts [1]. Several researchers have developed automated reasoning systems for diagrams. For example, Jamnik et al. [2] developed DIAMOND for automating diagrammatic proofs of arithmetic theorems. Barwise and Etchemendy [3] used blocks-world to teach and reason in first order logic with Hyperproof. Stapleton et al. [4] developed Edith for automated Euler diagram theorem proving. Urbas et al. [5] extended Edith to spider diagrams and developed Speedith. In these systems, mechanising reasoning with diagrams usually relies on methods of encoding diagrams as symbolic representation that can be easily processed with a rule-based program. Such methods rely on human experts to define the framework of diagram-to-symbol mapping and the set of rules to reason with the symbols. In this work, we developed a method using Deep artificial Neural Networks (DNN) to learn a continuous and vector-form neural representations of diagrams without any human input rules. With this method, diagrams can be encoded into neural representations, and reasoned about with subsequent neural networks or human-defined rules.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 390–398, 2018.
https://doi.org/10.1007/978-3-319-91376-6_36

Recently, DNNs have achieved human comparable performance in several tasks such as image recognition [6], natural language translation [7]. DNNs' success in tasks that humans are good at can be partly attributed to their biologically inspired architecture. Deep convolutional neural networks, a type of DNN applied for supervised image-based tasks, have both strong architectural and activity pattern similarity to visual cortex in the human brain [8]. This biological similarity motivates us to apply DNN to investigate diagrammatic reasoning. We developed Euler-Net, a type of DNN that performs syllogism reasoning with Euler diagrams. The DNN takes as input a number of Euler diagrams (premises), which show set relationships between sets contained in them. Euler-Net can generate a categorical conclusion (subset, intersection or disjoint) about the relationship between the sets with 99.5% accuracy. It can also learn using Generative Adversarial Network (GAN) to generate Euler diagrams that represent the set relationships without using any additional drawing tools. This enables Euler-Net to perform full diagrammatic inference, and shows that the learnt neural representations encodes essential information of the reasoning task.

Euler-Net is developed on, but not limited to, Euler diagram syllogism tasks. It can be applied to types of diagrams that are difficult to formalise into logic symbols, and can learn feature representations that capture essential information in the diagrams and subsequently analyse it. In our future work, we will adapt Euler-Net to a broader range of diagrams, and develop it into a useful tool for the diagram research community for investigating different types of diagrammatic representations.

In the rest of this paper we first give in Sect. 2 some background to neural nets and Euler diagrams. In Sect. 3 we present our deep neural net architecture and its use in Euler-Net to reason with Euler diagrams. Next, we evaluate our system in Sect. 4, and discuss these results and some future directions in Sect. 5. Finally we conclude in Sect. 6.

## 2 Background

### 2.1 Neural Networks

Artificial Neural Network (ANN) is an information processing paradigm that is inspired by biological nervous systems. A general ANN consists of layers of artificial neurons connected in a graph, most often in an acyclic directed form. A single artificial neuron has very limited computational capability. However, when many of them are connected together to form an ANN, very complex functions can be approximately represented. ANN can be optimised by back propagation algorithm [9], a way to back propagate errors from higher layers to lower layers in order to correct weights assigned to lower layer inputs.

Recently, a particular type of ANN, a convolutional neural network (CNN) has made breakthroughs in various AI tasks such as image classification [6]. CNN consists of convolutional layers that apply learnable filters on images to extract features, and pooling layers that summarise local information. A typical CNN

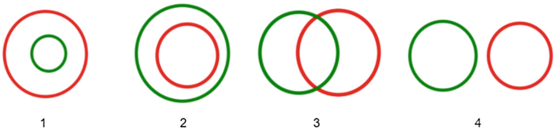


**Fig. 1.** Euler diagrams representing 4 possible relationships between non-empty sets A and B. (Color figure online)

consists of multiple stacks of convolutional and pooling layers. The layers learn to extract progressively more complex image features (from edges and contours to objects). A deep CNN can learn highly complex feature representations that are similar to neural responses in the human visual cortex [8].

## 2.2 Euler Diagrams

Euler diagrams [10] are a simple, yet effective diagrammatic representation for reasoning about set relationships. We will use a colour-coded modification of the Gergonne's system of Euler diagrams [11] for its simplicity and visual clarity. In this system, minimal regions are assumed to be non-empty (i.e., the Gergonne system of Euler diagrams assumes existential import), and shading is not used. We assign a distinct colour to each contour instead of alphabet labels to denote classes. Colour coding facilitates the training of neural network by reducing the need to associate Alphabet labels with circled regions. There can be four different relationships between two sets A and B, which are: (1) $A \supset B$, (2) $A \subset B$, (3) $A \cap B \neq \emptyset$ and (4) $A \cap B = \emptyset$.

While in theory the fifth relationship $A = B$ is also possible, we do not consider it in this work because colour-coded contours will completely overlap and thereby diminish visual clarity – we leave to explore this in the future. Figure 1 illustrates how these 4 different set relationships can be represented by 4 different categories of colour-coded Euler diagrams (two sets denoted by Red and Green).

Euler diagrams are very effective in representing syllogisms. A syllogism consists of two premises that entail a conclusion. Figure 2 illustrates how the colour-coded Euler diagrams represent the syllogism "All *Green* are *Red*, all *Blue* are *Green*, therefore all *Blue* are *Red*". In our task, we do not enforce the fixed-size contour constraint, which means that contours representing the same class can have varying sizes in different diagrams. As this Euler diagram system does not represent partial information, for certain premises there is not a single directly implied conclusion diagram, but several diagrams that are self-consistent [12] with the given premises. An example would be for premises "All B are A, some C are B", consistent conclusions include "some C are A" and "all C are A". We later show that neural networks can learn to reason with self-consistency and generate all conclusions that are consistent with the premises.

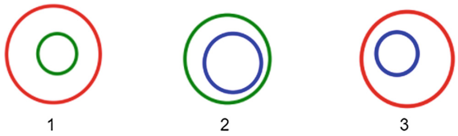


**Fig. 2.** Euler diagrams representing the syllogism "All *Green* are *Red*, all *Blue* are *Green*, therefore all *Blue* are *Red*". (Color figure online)

# 3 Neural Network Architecture

## 3.1 Euler-Net for Categorical Output

We built a system, called Euler-Net, which is a neural network trained to solve syllogisms represented with Euler diagrams. Euler-Net takes two Euler diagrams representing the premises of the syllogism as input, and outputs a categorical conclusion (subset, intersection or disjoint) for the syllogism. Figure 3 shows the architecture of Euler-Net. The first input diagram shows a relationship between a set *Red* and a set *Green*. The second diagram shows a relationship between sets *Green* and *Blue*. There are 4 possible categories for a categorical conclusion output from Euler-Net, as discussed in Sect. 2.2.

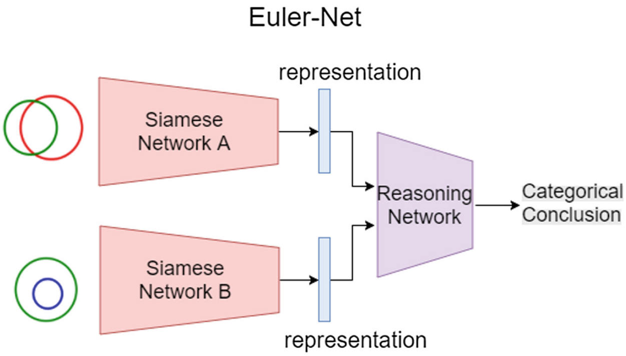


**Fig. 3.** Overview of the Euler-Net neural network architecture for Euler diagram syllogism reasoning. (Color figure online)

In the case of indeterminacy when no single logical conclusion can be drawn (e.g., All *Green* are *Red*, Some *Green* are *Blue*), the neural network outputs all conclusions that are consistent (not contradicting) with the premises. Neural networks are trained with pairs of premise diagrams and labels that encode correct conclusions into a binary (0 or 1) vector of length 4, with 4 positions in the vector representing each conclusion category as in Sect. 2.2. For example, vector [0100] encodes $Blue \subset Red$. While developed on the classical syllogisms, Euler-Net can be applied to diagram tasks with arbitrary number of contours in the diagram and arbitrary number of diagrams in the task. We also applied Euler-Net on tasks where there are 3 contours in each diagram. Namely, the first diagram contains *Red*, *Green* and *Blue* contours, while the second diagram contains *Green*, *Blue* and *Yellow* contours. The task is to infer consistent relationships between classes *Red* and *Yellow*. In Sect. 4 we show that for 3-contour task, the reasoning accuracy decrease is negligible.

Euler-Net is composed of two modules. The first module is a convolutional network that recognises the diagrams and encodes them into high-level neural feature representations. This network has a similar function to a visual cortex in the human brain [8], which transforms visual stimuli into neural code. The second module is a reasoning network that performs inference on the neural presentations of diagrams. This reasoning network extracts useful information from the neural representations in order to achieve accurate inference. The reasoning network consists of fully-connected layers that densely process the neural representations, and outputs the probability for each categorical conclusion. Euler-Net can be trained to minimise error rates in reasoning with standard Stochastic Gradient Descent (SGD) and a back-propagation algorithm [9]. Formally, the training objective is to minimise the loss function as in Eq. 1:

$$L(D,T) = -\sum_{(\mathbf{d},\mathbf{t})\in(D,T)} \sum_i t_i \log f(d) + (1-t_i)(1-log f(d)) \tag{1}$$

where $D$ are premise diagrams, $T$ are labels, $(d,t)$ is a training sample of the problem set, $t_i$ is the $i^{th}$ element in the label vector, and $f(d)$ represents Euler-Net as a function of $d$.

### 3.2 Euler-Net for Diagram Generation

Instead of generating a categorical conclusion, Euler-Net can also generate diagrammatic conclusion of a syllogism, such as diagram 3 in Fig. 2. This allows Euler-Net to perform complete diagrammatic inference. Diagrams can be generated from the neural representations of the syllogism problem without any human intervention or established drawing tools. This is accomplished by concatenating an image generator network to Euler-Net. Figure 4 illustrates the architecture of Euler-Net for diagram generation. This generator network uses latent neural code vector extracted from the last layer of Euler-Net to generate Euler diagrams that are consistent with the given premises. The latent code vector encodes consistent conclusions for the reasoning task. The generator network then transform this neural representation to an Euler diagram consistent with the given premises.

The generator network is trained with Generative Adversarial Network (GAN) [13] training objective, which recently became popular for generating high definition and sharp images. GAN consists of a generator network and a discriminator network that are jointly trained in a minimax game. The generator tries to generate images as real and accurate as possible, while the discriminator

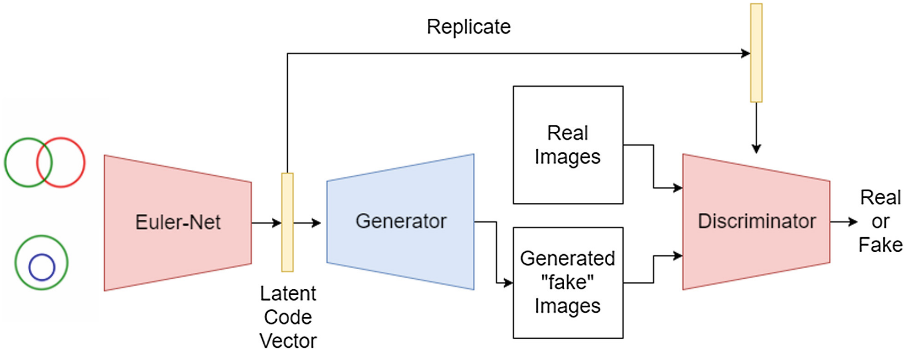


**Fig. 4.** Overview of diagram generation module for Euler-Net.

tries to distinguish between the generated and the correct images. The GAN training objective can be mathematically formulated as in Eq. 2:

$$\min_{G} \max_{D} V(G,D) = \mathbf{E}_{x \sim p_{data}(x)}[log D(x)] + \mathbf{E}_{z \sim p_z(z)}[log(1 - D(G(z)))] \quad (2)$$

where $G$ is the generator, $D$ is the discriminator, $x$ is a correct data sample, and $z$ is the latent code vector. This can be viewed as a minimax game between $G$, which tries to minimise the objective, and $D$ which tries to maximise it. During training, the parameters of the generator and the discriminator are updated alternatively to converge towards a dynamic equilibrium. The generator converges after 50000 iterations, and is able to generate an accurate and clear Euler diagram conclusion consistent with the given premises.

## 4 Evaluation

Euler-Net is trained with syllogism problems generated from an Euler diagram syllogism task generator. This generator firstly generates two random logical relationships for the first two premises, and then generates two Euler diagrams representing the two logical relationships with random size and position, as long as the logical relationships are not violated. Subsequently, the task generator generates consistent conclusions from a manually constructed truth table that maps any two premises to a set of consistent conclusions. For each consistent conclusion, corresponding Euler diagrams are also generated with random size and position. In total, we generated 96000 Euler syllogism reasoning problems for neural network training. For the 3-contour dataset, the diagrams and conclusions are generated in the same fashion. The truth table is larger as each premise now contains relationships between 3 classes, giving $4^3$ possible cases.

During training, we divide the dataset into training set, validation set and test set with a split ratio of $8:1:1$. We trained Euler-Net with the training dataset, tuned its performance with reported scores on the validation dataset, and finally, we evaluated the final performance on the test dataset. We report

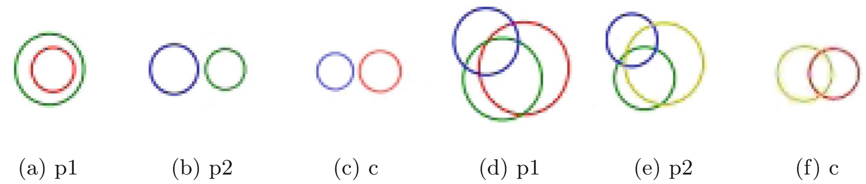

(a) p1 (b) p2 (c) c (d) p1 (e) p2 (f) c

**Fig. 5.** Two examples of the Euler-Net generated diagrams. p1 and p2 are premises and c are Euler-Net generated conclusion diagrams.

here the percentage accuracy which is defined as the number of syllogism problems correctly solved over the total number of problems. Euler-Net is able to achieve a nearly perfect accuracy of 99.5% on the 2-contour Euler syllogism tasks, and 99.4% on 3-contour task. In order to understand the performance results further, we separate test results for conclusive syllogism groups (a single logical conclusion) and inconclusive groups (multiple consistent conclusions). We found that Euler-Net achieves 100% accuracy for conclusive groups, which could indicate that conclusive syllogisms are relatively simpler than the inconclusive ones.

Euler-Net with added diagram generator can create high quality Euler diagrams from neural representations mostly without image artefacts. Figure 5 shows examples of Euler conclusion diagrams generated by Euler-Net for 2-contour and 3-contour task. This shows that Euler-Net learns neural representations that encode essential information of the syllogism reasoning problem. Such neural representations can be interpreted by a diagram generator network to create a diagrammatic conclusion.

## 5 Discussion

While we showed that DNN can perform diagrammatic reasoning and learn useful representations on the relatively simple Gergonne's system of Euler diagram syllogism solving, DNN is not limited to a particular type of diagram. DNN provides a universal method for encoding all types of diagrams into neural codes that can be subsequently analysed. While simple diagrams like Euler diagrams can be conveniently formalised symbolically into sets of zones, labels and shadings, there are many types of diagrams that are more difficult to formalise. DNN can be applied to learn representations of such diagrams, and thus enable us to analyse aspects of those representations, such as for example, the difficulty level of the question, the amount of redundant information in the diagram, and how the core logic can be efficiently encoded with DNN in the sequence of diagrams. We will apply DNN to such more complex diagrams in our future work.

Euler-Net's learning capacity can be increased for more complex diagrams than Euler diagrams with 3 contours by increasing architecture complexity, and including recent techniques such as residual network connections. While in theory for n-set Euler diagram, quadratic amounts of nodes are required in the network, most of the nodes can be designed as a reusable module for simple operations, similarly to how we use the CNN weights for input diagrams. We can also introduce a recent technique called "neural network attention", which allows a neural network to reuse weights for learning simpler features such as 2-set relations – this would greatly improve scalability.

While Euler-Net achieves 99.5% accuracy, it is still not on-par with logical symbolic reasoner which is 100% accurate. However, conceptually, logical symbolic reasoners only reason with the symbols that represent diagrams, while Euler-Net reasons directly with the raw diagram input. The strength of DNN is its ability to capture feature representations for any type of diagram and its robustness to noise. In our future work we will extend Euler-Net to diagrams that are drawn with noise or even missing parts, and where a symbolic reasoner may fail because of difficulty of transforming noisy diagrams into symbols. We will also develop Euler-Net on a full set of modern Euler-typed diagrams with shading and non-existential import, including concept diagrams and spider diagrams. We will develop Euler-Net as a neural-network-counterpart of modern diagram theorem provers like Speedith [5].

## 6 Conclusion

We developed Euler-Net, a deep neural network that can learn to perform diagrammatic reasoning about Euler diagram syllogism tasks. We illustrated that neural representations learnt by Euler-Net encode essential information about the input diagrams and the reasoning task. Euler-Net, while developed on Euler diagrams, is not constrained to a specific category of diagrams. Euler-Net can be readily adapted to other more complex diagrams where information extraction and formalisation is more difficult. We believe that deep neural networks can be developed into a useful tool for the diagram research community.

## References

1. Simon, H.A., Larkin, J.H.: Why a diagram is (sometimes) worth 10,000 words. Models Thought **2** (1987)
2. Jamnik, M., Bundy, A., Green, I.: On automating diagrammatic proofs of arithmetic arguments. J. Log. Lang. Inf. **8**(3), 297–321 (1999)
3. Barwise, J., Etchemendy, J.: Hyperproof. CSLI Press, California (1994)
4. Stapleton, G., Masthoff, J., Flower, J., Fish, A., Southern, J.: Automated theorem proving in Euler diagram systems. J. Autom. Reason. **39**(4), 431–470 (2007)
5. Urbas, M., Jamnik, M., Stapleton, G.: Speedith: a reasoner for spider diagrams. J. Log. Lang. Inf. **24**(4), 487–540 (2015)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
7. Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al.: Google's neural machine translation system: bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144 (2016)

8. Yamins, D.L., DiCarlo, J.J.: Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. **19**(3), 356–365 (2016)
9. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
10. Hammer, E., Shin, S.J.: Eulers visual logic. Hist. Philos. Log. **19**(1), 1–29 (1998)
11. Gergonne, J.D.: Essai de dialectique rationelle. Annuales de Mathematiques pures et appliqukes **7**, 189–228 (1817)
12. Shimojima, A.: Semantic Properties of Diagrams and Their Cognitive Potentials. Center for the Study of Language & Information, California (2015)
13. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)

# Sequent Calculus for Euler Diagrams

Sven Linker(✉)

University of Liverpool, Liverpool, UK
s.linker@liverpool.ac.uk

**Abstract.** Proof systems play a major role in the formal study of diagrammatic logical systems. Typically, the style of inference is not directly comparable to traditional sentential systems, to study the diagrammatic aspects of inference. In this work, we present a proof system for Euler diagrams with shading in the style of sequent calculus. We prove it to be sound and complete. Furthermore we outline how this system can be extended to incorporate heterogeneous logical descriptions. Finally, we explain how small changes allow for reasoning with intuitionistic logic.

**Keywords:** Euler diagrams · Proof systems · Heterogeneous reasoning

## 1 Introduction

Starting from the early work on formal diagrammatic systems, the analysis of proof systems has played a major role. For example, in the seminal work of Shin [10], she developed a proof system for each system of Venn-diagrams she defined, and proved each to be sound and complete. Unsurprisingly, comparing typical rules for Euler and Venn-diagrams with sentential rules is hard. This is mainly due to two reasons. On the one hand, the former rules are inherently diagrammatic in nature and are often not directly comparable to sentential rules. For example, *introducing* a new contour into an Euler diagram is defined such that the logical information in the diagram is not affected. That is, from a logical perspective, the original diagram and the changed one are equivalent. While such changes are at least unusual for sentential transformations, diagrammatic proof systems make considerable use of equivalent transformations. On the other hand, proofs for Euler diagrams or Spider diagrams are defined as a linear progression from the assumptions to the conclusion [6,10], while sentential proofs are most of the time defined as proof-trees, where an application of a rule may split the current proof state into branches, e.g., in systems of natural deduction or sequent calculus [9]. To the best of our knowledge, the only direct comparison between diagrammatic inference systems and sentential reasoning styles was conducted by Mineshima et al. [8]. They analysed proof systems for two diagrammatic languages: Euler diagrams without shading and Venn-Diagrams, and showed, how the former relates to natural deduction, and the latter to resolution calculus.

---

This work was supported by EPSRC Research Programme EP/N007565/1 *Science of Sensor Systems Software*.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 399–407, 2018.
https://doi.org/10.1007/978-3-319-91376-6_37

In this work, we present a proof system for Euler diagrams with shading in the style of sequent calculus [5]. We prove this system to be sound and complete. Furthermore, we explain how simple amendments allow us to create a system for a heterogeneous language of Euler diagrams and propositional logic.

This paper is structured as follows. In Sect. 2, we give a short definition of Euler diagrams and the semantics we use. Section 3 contains the definition of the calculus and the proofs for soundness and completeness, while we discuss extensions and relations to other systems and conclude in Sect. 4.

## 2 Euler Diagrams

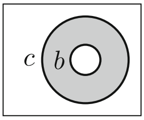


**Fig. 1.** Euler diagram

An Euler diagram consists of a set of *contours*, dividing the space enclosed by a bounding rectangle into different, possibly shaded zones (see Fig. 1 for an example). Traditionally, each contour represents a set, and the diagram restricts the possible relations between these sets. We take a slightly different approach: contours represent propositional variables, taken from a countably infinite set $\mathsf{Vars}$. A *zone* for a finite set of contours $L \subset \mathsf{Vars}$ is a tuple $(\mathsf{in}, \mathsf{out})$, where $\mathsf{in}$ and $\mathsf{out}$ are disjoint subsets of $L$ such that $\mathsf{in} \cup \mathsf{out} = L$. The set of all zones for a given set of contours is denoted by $\mathsf{Venn}(L)$.

**Definition 1 (Abstract Syntax).** *A* unitary *Euler diagram is a tuple $d = (L, Z, Z^*)$, where $Z$ and $Z^*$ are sets of zones for $L$ such that $Z^* \subseteq Z$, $(\emptyset, L) \in Z$, and for each $c \in L$, there is a zone $(\mathsf{in}, \mathsf{out}) \in Z$, such that $c \in \mathsf{in}$. The set $Z$ denotes the* visible *and $Z^*$ the* shaded zones *of $d$. For a unitary diagram $d$, we will also refer to the set of its* missing zones $\mathsf{MZ}(d) = \mathsf{Venn}(L) \setminus Z$. *The syntax of Euler diagrams is then given as $D ::= d \mid D \rightarrow D$, where $d$ is unitary. Euler diagrams of the form $D_1 \rightarrow D_2$ are* compound.

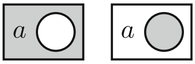


**Fig. 2.** Literals

We allow the diagrams $\top = (\emptyset, \{(\emptyset, \emptyset)\}, \emptyset)$ and $\bot = (\emptyset, \{(\emptyset, \emptyset)\}, \{(\emptyset, \emptyset)\})$. We define negation by $\neg D \equiv D \rightarrow \bot$ and the missing connectives $D_1 \vee D_2 \equiv \neg D_1 \rightarrow D_2$, etc. A *literal* is a unitary diagram for a single contour, with exactly one shaded zone. If the zone $(\emptyset, \{c\})$ is shaded in a literal, then we call it *positive*, otherwise it is *negative* (see Fig. 2). Observe that our notion of literals slightly deviates from the original definition of Stapleton and Masthoff [11].

**Definition 2 (Semantics).** *A* valuation *is a function $\nu\colon \mathsf{Vars} \rightarrow \mathbb{B}$, where $\mathbb{B} = \{\mathsf{tt}, \mathsf{ff}\}$ is the set of Boolean values. We denote the set of all valuations by* $\mathsf{Vals}$. *Let $z = (\mathsf{in}, \mathsf{out})$ be a zone. The* semantics *of $z$ is a subset of* $\mathsf{Vals}$, *given by $[\![z]\!] = \{\nu \mid \forall c \in \mathsf{in}\colon \nu(c) = \mathsf{tt} \text{ and } \forall c \in \mathsf{out}\colon \nu(c) = \mathsf{ff}\}$. The semantics of Euler diagrams is then $[\![d]\!] = \bigcup_{z \in Z \setminus Z^*} [\![z]\!]$ and $[\![D_1 \rightarrow D_2]\!] = (\mathsf{Vals} \setminus [\![D_1]\!]) \cup [\![D_2]\!]$, where $d$ is unitary and $D_1$, $D_2$ are arbitrary Euler diagrams. If $[\![D]\!] = \mathsf{Vals}$, then we call $D$* valid, *denoted by $\models D$. Otherwise, $D$ is* falsifiable.

Note that $[\![\top]\!] = \mathsf{Vals}$ and $[\![\bot]\!] = \emptyset$, as well as $[\![D_1 \vee D_2]\!] = [\![D_1]\!] \cup [\![D_2]\!]$ and $[\![D_1 \wedge D_2]\!] = [\![D_1]\!] \cap [\![D_2]\!]$. Furthermore, the semantics of a positive literal for the contour $c$ consists of the valuations with $\nu(c) = \mathsf{tt}$.

**Definition 3 (Adjacent Zone).** *Let* $z = (\mathsf{in}, \mathsf{out})$ *be a zone for the contours in* $L$ *and* $c \in L$*. The* zone adjacent to $z$ *at* $c$*, denoted by* $\overline{z}^c$ *is* $(\mathsf{in} \cup \{c\}, \mathsf{out} \setminus \{c\})$*, if* $c \in \mathsf{out}$ *and* $(\mathsf{in} \setminus \{c\}, \mathsf{out} \cup \{c\})$ *if* $c \in \mathsf{in}$*.*

Now we can define a way to remove contours from a unitary diagram $d$.

**Definition 4 (Reduction).** *Let* $d = (L, Z, Z^*)$ *be a unitary Euler diagram and* $c \in L$*. The* reduction *of a zone* $z = (\mathsf{in}, \mathsf{out})$ *is defined by* $z \backslash c = (\mathsf{in} \backslash \{c\}, \mathsf{out} \backslash \{c\})$*. The* reduction *of* $d$ *by* $c$ *is defined as* $d \setminus c = (L \setminus \{c\}, Z \setminus c, Z^* \setminus c)$*, where*

$$Z \setminus c = \{z \setminus c \mid z \in Z\}$$
$$Z^* \setminus c = \{z \setminus c \mid z \in Z^* \text{ and } \overline{z}^c \in Z^* \cup \mathsf{MZ}(d)\}$$

*That is, we remove the contour* $c$ *from all zones and only shade the reduction of a shaded zone* $z$*, if its adjacent zone at* $c$ *is shaded or missing.*

If each shaded or missing zone in a diagram $d$ has a shaded or missing adjacent zone, then the conjunction of the reduction of $d$ by each of its contours preserves the semantic information. That is, we can distribute the information contained in $d$ among simpler diagrams.

**Lemma 1.** *Let* $d = (L, Z, Z^*)$*, where for each* $z \in Z^* \cup \mathsf{MZ}(d)$*, there is a contour* $\ell \in L$ *such that* $\overline{z}^\ell \in Z^* \cup \mathsf{MZ}(d)$*. Then* $[\![d]\!] = \bigcap_{c \in L} [\![d \setminus c]\!]$

*Proof.* For each $c \in L$, we have $[\![d]\!] \subseteq [\![d \setminus c]\!]$. Hence the direction from left to right is immediate. Now let $d = (L, Z, Z^*)$ and $\nu$ be such that $\nu \in [\![d \setminus c]\!]$ for all $c \in L$. Hence, for each $c$, there is a $z_c \in Z$, such that $\nu \in [\![z_c \setminus c]\!]$. Now we have to show that in fact there is a *single* zone $z \in Z$, such that $\nu \in [\![z \setminus c]\!]$ for all $c$. Observe that there are two zones $z_{\mathsf{tt}}, z_{\mathsf{ff}} \in \mathsf{Venn}(d)$ such that $\nu \in [\![z_{\mathsf{tt}} \setminus c]\!]$ and $\nu \in [\![z_{\mathsf{ff}} \setminus c]\!]$, whose only difference is that $c$ is in the in-set of $z_{\mathsf{tt}}$ and in the out-set of $z_{\mathsf{ff}}$. Now, assume $\nu(c) = \mathsf{tt}$, hence $\nu \in [\![z_{\mathsf{tt}}]\!]$. If $\nu \notin [\![d]\!]$, this means that $z_{\mathsf{tt}} \in Z^* \cup \mathsf{MZ}(d)$. By assumption, there is a contour $\ell$ such that $\overline{z_{\mathsf{tt}}}^\ell \in Z^* \cup \mathsf{MZ}(d)$. In the reduction of $d$ by $\ell$, this means that $z_{\mathsf{tt}}$ is either shaded or missing as well, and hence $\nu \notin [\![d \setminus \ell]\!]$, which contradicts the assumption on $\nu$. Hence $\nu \in [\![d]\!]$. The case for $\nu(c) = \mathsf{ff}$ is similar. □

## 3 Sequent Calculus

Sequent calculus, as defined by Gentzen [5] is closely related to natural deduction. It is based on *sequents*, which are decomposed by rule applications.

$$\frac{\Gamma, D, E \Rightarrow \Delta}{\Gamma, D \wedge E \Rightarrow \Delta}\,\mathrm{L}\wedge \qquad \frac{\Gamma, D \Rightarrow \Delta \quad \Gamma, E \Rightarrow \Delta}{\Gamma, D \vee E \Rightarrow \Delta}\,\mathrm{L}\vee \qquad \frac{\Gamma \Rightarrow \Delta, D \quad \Gamma, E \Rightarrow \Delta}{\Gamma, D \rightarrow E \Rightarrow \Delta}\,\mathrm{L}\rightarrow$$

$$\frac{\Gamma \Rightarrow \Delta, D \quad \Gamma \Rightarrow \Delta, E}{\Gamma \Rightarrow \Delta, D \wedge E}\,\mathrm{R}\wedge \qquad \frac{\Gamma \Rightarrow \Delta, D, E}{\Gamma \Rightarrow \Delta, D \vee E}\,\mathrm{R}\vee \qquad \frac{\Gamma, D \Rightarrow \Delta, E}{\Gamma \Rightarrow \Delta, D \rightarrow E}\,\mathrm{R}\rightarrow$$

**Fig. 3.** Proof rules for Boolean operators

**Definition 5 (Sequent).** *A* sequent $\Gamma \Rightarrow \Delta$ *consists of two multisets* $\Gamma$ *and* $\Delta$ *of Euler diagrams. The multiset* $\Gamma$ *is called the* antecedent *and* $\Delta$ *the* succedent.

*If* $\Gamma$ *(*$\Delta$*) is the empty multiset, we write* $\Rightarrow \Delta$ *(*$\Gamma \Rightarrow$*, respectively). If a sequent is of the form* $\Gamma, l \Rightarrow \Delta, l$*, where* $l$ *is a positive literal, or* $\Gamma, \bot \Rightarrow \Delta$*, or* $\Gamma \Rightarrow \Delta, \top$ *then it is called an* axiom. *A sequent* $D_1, \ldots, D_k \Rightarrow E_1, \ldots, E_n$ *is equivalent to* $(D_1 \wedge \cdots \wedge D_k) \rightarrow (E_1 \vee \cdots \vee E_n)$*. The notions of validity and falsifiability carry over from the semantics of Euler diagrams.*

A *deduction* for a sequent $\Gamma \Rightarrow \Delta$ is a tree, where the root is labelled by $\Gamma \Rightarrow \Delta$, and the children of each node are labelled according to the rules defined below. A deduction where each leaf is labelled with an axiom is called a *pf* for $\Gamma \Rightarrow \Delta$. We denote the existence of a proof for $\Gamma \Rightarrow \Delta$ by $\vdash \Gamma \Rightarrow \Delta$. Intuitively, the prover tries to refute the sequent, i.e., she tries to find a valuation that satisfies all diagrams in the antecedent and falsifies every diagram in the succedent. If all possible ways to find such a valuation fail, i.e., each branch ends with an axiomatic sequent, then the diagram is valid. For proof search, it is beneficial to apply the rules backwards, that is from bottom to top.

**Lemma 2.** *A sequent containing only positive literals is valid iff it is an axiom.*

*Proof.* The right to left direction is immediate. Now let $d_1, \ldots, d_k \Rightarrow e_1, \ldots, e_n$ be valid, where each $d_i$ and $e_j$ is a positive literal, and assume it is not an axiom. Hence, for no $i$ and $j$, we have that $d_i = e_j$. Then the valuation $\nu$ with $\nu(d_i) = \mathsf{tt}$ and $\nu(e_j) = \mathsf{ff}$ falsifies the sequent, which contradicts our assumption. $\square$

The rules to treat compound diagrams, as shown in Fig. 3, are directly taken from sequent calculus for propositional logic and are sound [9].

**Lemma 3 (Soundness).** *The rules for Boolean operators are sound.*

Let $d = (L, Z, Z^*)$ with $|Z^*| > 1$, and let $d_i = (L, Z, Z_i^*)$, for $i \in \{1, 2\}$, such that $Z^* = Z_1^* \cup Z_2^*$. Then the rules $L\mathsf{s}$ and $R\mathsf{s}$ shown in Fig. 4a *separate* $d$. These rules are closely related to the *Combine* equivalence rule for Spider diagrams [6].

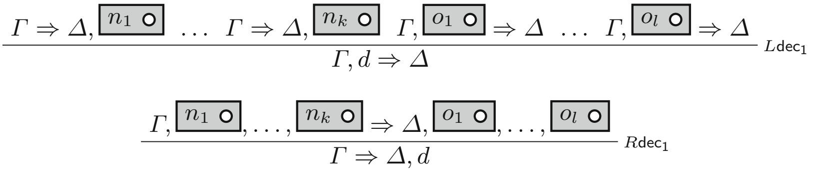
For $d = (L, Z, Z^*)$ with $|\mathsf{MZ}(d)| > 0$ and $z \in \mathsf{MZ}(d)$, let $d^z = (L, Z \cup \{z\}, Z^* \cup \{z\})$. The rules $L\mathsf{MZ}$ and $R\mathsf{MZ}$ in Fig. 4b *introduce the missing zone* $z$.

Now let $d = (L, Z, Z^*)$, where for each $z \in Z^* \cup \mathsf{MZ}(d)$ there is a contour $\ell \in L$, such that $\overline{z}^{\ell} \in Z^* \cup \mathsf{MZ}(d)$. Let $L = \{c_1, \ldots, c_k\}$. Then we can *reduce* $d$ according to the rules $L\mathsf{r}$ and $R\mathsf{r}$ shown in Fig. 4c.

$$\dfrac{\Gamma, d_1, d_2 \Rightarrow \Delta}{\Gamma, d \Rightarrow \Delta}\ Ls \qquad \dfrac{\Gamma, d^z \Rightarrow \Delta}{\Gamma, d \Rightarrow \Delta}\ L\mathsf{MZ} \qquad \dfrac{\Gamma, d \setminus c_1, \ldots, d \setminus c_k \Rightarrow \Delta}{\Gamma, d \Rightarrow \Delta}\ Lr$$

$$\dfrac{\Gamma \Rightarrow \Delta, d_1 \quad \Gamma \Rightarrow \Delta, d_2}{\Gamma \Rightarrow \Delta, d}\ Rs \qquad \dfrac{\Gamma \Rightarrow \Delta, d^z}{\Gamma \Rightarrow \Delta, d}\ R\mathsf{MZ} \qquad \dfrac{\Gamma \Rightarrow \Delta, d \setminus c_1 \ \ldots \ \Gamma \Rightarrow \Delta, d \setminus c_k}{\Gamma \Rightarrow \Delta, d}\ Rr$$

(a) (b) (c)



(d)

**Fig. 4.** Proof rules to decompose unitary diagrams

Finally, let $d = (L, Z, Z^*)$, where $d$ is not a positive literal or $\bot$, and either $|Z^*| = 1$ and $|\mathsf{MZ}(d)| = 0$ or $|\mathsf{MZ}(d)| = 1$ and $|Z^*| = 0$. Let $z = (\mathsf{in}, \mathsf{out})$ be the corresponding shaded or missing zone, where $\mathsf{in} = \{n_1, \ldots, n_k\}$ and $\mathsf{out} = \{o_1, \ldots, o_l\}$. Then the rules $L\mathsf{dec}_1$ and $R\mathsf{dec}_1$ (see Fig. 4d) decompose $d$ into positive literals.

An example of a proof is shown in Fig. 5. In the applications of $L\mathsf{dec}_1$ and $R\mathsf{dec}_1$, the diagram denoting the disjointness of $u$ and $w$ is decomposed on the left side (right side, resp.) of the sequent. The application of $L\mathsf{r}$ is possible, since for each shaded or missing zone $z$, there is a contour $c$ such that $\overline{z}^c$ is also shaded or missing. E.g., consider $z = (\{u\}, \{v, w\})$. Then $\overline{z}^w = (\{u, w\}, \{v\})$ is missing. Hence, in the reduction of the diagram by $w$, the zone $(\{u\}, \{v\})$ is also shaded. That is, we can decompose a complex diagram into simpler diagrams, whose conjunction comprises the same information as the original.

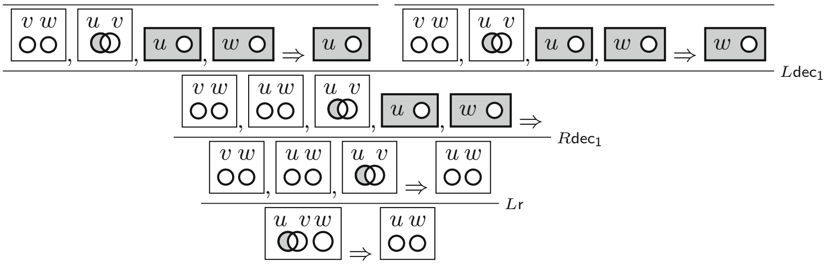


**Fig. 5.** Example of a deduction

**Lemma 4.** *The conclusion of each rule in Fig. 4 is falsifiable, if and only if, at least one of its premises is falsifiable.*

*Proof.* First we consider $L\mathsf{s}$. Let $d = (L, Z, Z^*)$, where $|Z^*| > 1$, and $d_1$, $d_2$ be as required for an application of $L\mathsf{s}$. Furthermore, let $\nu$ be a valuation that falsifies $\Gamma, d \Rightarrow \Delta$, i.e., $\nu$ satisfies $\Gamma$ and $d$, and falsifies $\Delta$. Since $Z^* = Z_1^* \cup Z_2^*$, we have $[\![d]\!] = [\![d_1]\!] \cap [\![d_2]\!]$. Hence $\nu$ falsifies $\Gamma, d \Rightarrow \Delta$ if and only if $\nu$ falsifies $\Gamma, d_1, d_2 \Rightarrow \Delta$. For $R\mathsf{s}$, let $\nu$ falsify $\Gamma \Rightarrow \Delta, d$. That is, $\nu$ falsifies $d$. Since $[\![d]\!] = [\![d_1]\!] \cap [\![d_2]\!]$, this is equivalent to $\nu$ falsifying at least one of $d_1$ and $d_2$.

The rules $L\mathsf{r}$ and $R\mathsf{r}$ can be proven sound similarly, due to Lemma 1.

For $L\mathsf{dec}_1$ and $R\mathsf{dec}_1$, let $z = (\mathsf{in}, \mathsf{out})$ be the single shaded zone in $d$ (the case for $z$ being missing is similar). Now consider $L\mathsf{dec}_1$. Let $\nu$ be a valuation that falsifies $\Gamma, d \Rightarrow \Delta$. Hence, $\nu \in [\![d]\!]$. That is, either $\nu(n) = \mathsf{ff}$ for at least one $n \in \mathsf{in}$, or $\nu(o) = \mathsf{tt}$ for at least one $o \in \mathsf{out}$. Assume that $\nu(n_i) = \mathsf{ff}$ (the other case is similar). This is equivalent to $\nu$ falsifying $\Gamma \Rightarrow \Delta, \boxed{n_i \ \circ}$. Consider $R\mathsf{dec}_1$. If $\nu$ falsifies $\Gamma \Rightarrow \Delta, d$, then $\nu \notin [\![d]\!]$. Since $[\![d]\!] = \mathsf{Vals} \setminus [\![z]\!]$, we have $\nu \in [\![z]\!]$. That is, $\nu(n) = \mathsf{tt}$ and $\nu(o) = \mathsf{ff}$ for all $n \in \mathsf{in}$ and $o \in \mathsf{out}$. Hence $\nu$ falsifying the premiss of $R\mathsf{dec}_1$ is equivalent to $\nu$ falsifying $\Gamma \Rightarrow \Delta, d$.

$L\mathsf{MZ}$ and $R\mathsf{MZ}$ are sound, since missing and shaded zones are equivalent. □

From Lemmas 3 and 4, we immediately get the necessary soundness theorem.

**Theorem 1 (Soundness).** $\vdash \Gamma \Rightarrow \Delta$ *implies* $\models \Gamma \Rightarrow \Delta$.

*Proof.* By induction on the length of proofs, using Lemmas 3 and 4. □

To prove completeness, we need to show that each diagram can be decomposed into positive literals. That is, each deduction can be maximised until only positive literals remain. Note that the example in Fig. 5 is not maximal.

**Lemma 5.** *Every deduction for a sequent* $\Gamma \Rightarrow \Delta$ *can be extended to a maximal deduction, where all diagrams in each leaf are either positive literals,* $\bot$ *or* $\top$.

*Proof.* Assume we have a deduction for $\Gamma \Rightarrow \Delta$, where one of the leaves contains a diagram $D$, which is not a literal. If $D$ is compound, we use the rules for Boolean operators to decompose $D$, until we reach a sequent where $D$ is reduced to a set of unitary diagrams (possibly on both the left and the right side of the sequent). Now, let $d$ be such a unitary diagram. If $d$ contains only one shaded or missing zone, then depending on the side on which $d$ appears, we can apply $L\mathsf{dec}_1$ or $R\mathsf{dec}_1$ to decompose $d$ to literals. Otherwise, we have to distinguish two cases. If $d$ contains more than one missing zone, we can apply $L\mathsf{MZ}$ or $R\mathsf{MZ}$ to change them to shaded zones. If $d$ contains more than one shaded zone, we can repeatedly apply $L\mathsf{s}$ or $R\mathsf{s}$ to separate $d$ to diagrams which only contain a single shaded zone. Finally, if $d$ does not contain any shaded or missing zones, we can reduce it by using $L\mathsf{r}$ and $R\mathsf{r}$. We can repeat these steps for every diagram in the leaves for the derivation of $\Gamma \Rightarrow \Delta$. Since each step reduces the number of operators or of missing or shaded zones, this yields a maximal derivation. □

**Theorem 2 (Completeness).** $\models \Gamma \Rightarrow \Delta$ *implies* $\vdash \Gamma \Rightarrow \Delta$.

*Proof.* Assume $\models \Gamma \Rightarrow \Delta$. By Lemma 5, we can create a maximal derivation for $\Gamma \Rightarrow \Delta$. Since $\Gamma \Rightarrow \Delta$ is valid, the premises constructed in each step are valid as well, due to Lemmas 3 and 4. Hence the leaves of the deduction tree are valid, and the only valid leaves are axioms by Lemma 2. Accordingly, $\vdash \Gamma \Rightarrow \Delta$. $\square$

## 4 Discussion

In this section, we compare our calculus with existing proof systems for Euler diagrams and discuss its properties, implications and possible extensions.

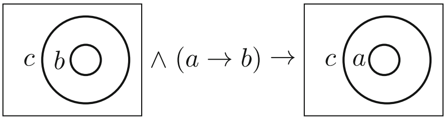


**Fig. 6.** Heterogeneous Euler diagrams

Burton et al. analysed strategies for completeness proofs of diagrammatic languages [3]. They emphasise that usually the strategy how to prove an Euler diagram $E$ from the assumptions $D_1, \ldots, D_n$ is to first create a maximal diagram $D_{max}$ incorporating the information from all of the $D_i$. Then, all information that is not part of $E$ will be removed from $D_{max}$. We do not apply this strategy. Instead, the rules decompose the diagrams, with the only exception being the rules to introduce missing zones. This is due to the similarity of our calculus to typical sentential calculi.

The proof system presented in this paper is related to both systems presented by Mineshima et al. [8]. It is oriented towards refutations, like the resolution calculus for Venn-diagrams, but also contains rules for the connectives and diagrammatic elements, like natural deduction for Euler diagrams. However, our language comprises both Venn-diagrams and Euler diagrams without shading.

We can extend our calculus to facilitate heterogeneous sequents in a rather simple way. We can allow compound diagrams to be mixed with propositional formulas, as for example shown in Fig. 6. The rules for Boolean operators can then be directly applied to propositional formulas. The only extension we need to incorporate into the calculus are heterogeneous axioms.

$\Gamma, a \Rightarrow \Delta,$ 

$\Gamma,$ 

$\Rightarrow \Delta, a$

This system then allows us to reason about heterogeneous Euler diagrams. However, it is hardly a heterogeneous *reasoning* system in the sense of Barwise and Etchemendy [2], since it does not include rules to transfer information from one representation into the other.

Furthermore, it is simple to amend our calculus to represent intuitionistic logic instead of classical propositional logic. To that end, we restrict the succedent of sequents to be a single formula, and change the Boolean rules accordingly[1]. For most of the diagrammatic rules, this change is sufficient as well, the only exceptions are $L\mathsf{dec}_1$ and $R\mathsf{dec}_1$. However, we can change these rules as follows.

---

[1] Compare with the textbook by Negri et al. [9].

$$\frac{\Gamma, d \Rightarrow \boxed{n_1 \circ} \quad \dots \quad \Gamma, d \Rightarrow \boxed{n_k \circ} \quad \Gamma, \boxed{o_1 \circ} \Rightarrow D \quad \dots \quad \Gamma, \boxed{o_l \circ} \Rightarrow D}{\Gamma, d \Rightarrow D}\ L\mathsf{dec}_1{}^I$$

$$\frac{\Gamma, \boxed{n_1 \circ}, \dots, \boxed{n_k \circ} \Rightarrow \boxed{o_i \circ}}{\Gamma \Rightarrow d}\ R\mathsf{dec}_1{}^I$$

That is, in $L\mathsf{dec}_1{}^I$, we keep the diagram in the antecedent for the branches with the new literals in the succedent, while we omit it in the branches, where we add literals to the antecedent. For $R\mathsf{dec}_1{}^I$, we choose a single occurrence of a literal to keep in the succedent. These changes are similar to the changes for the Boolean operators. Observe that the semantics presented in Sect. 2 is no longer suited for this proof system. We would have to define a semantics based on intuitionistic models, for example Heyting algebras. However, how such a semantics should look like is not obvious. It would be interesting to study the connection of this proof system to traditional proof systems for Euler diagrams, since the graphical notations for intuitionistic logic are sparse. Notable exceptions are the work of de Freitas and Viana [4], defining a graphical calculus for relational reasoning, and Alves et al. [1], in which they present a visualisation of intuitionistic proofs. In a similar way, we could try to change the system to reflect substructural logics, i.e., logics for which the structural rules of weakening, contraction and/or permutation do not hold[2]. However, in these logics, new operators arise and would have to be reflected in the diagrams as well. Such a radical change is not part of this paper, and left as future work. Of course, classical diagrammatic systems are possible ways to extend our calculus as well. A natural next step is an extension to treat Spider diagrams or Constraint diagrams.

A sequent calculus style proof system is suited for automatic proof search. Hence, an implementation into the theorem prover Speedith [12] is obvious future work, since it already supports backward reasoning and several proof branches. Furthermore, extending the tactics within Speedith [7] to our calculus would allow us to delay the application of rules creating new branches in the proof.

## References

1. Alves, S., Fernández, M., Mackie, I.: A new graphical calculus of proofs. In: Echahed, R. (ed.) TERMGRAPH 2011, vol. 48, pp. 69–84 (2011)
2. Barwise, J., Etchemendy, J.: Heterogeneous logic. In: Logical Reasoning with Diagrams, pp. 179–200. Oxford University Press Inc. (1996)
3. Burton, J., Stapleton, G., Howse, J.: Completeness proof strategies for Euler diagram logics. In: Euler Diagrams 2012, vol. 854, pp. 2–16. CEUR (2012)
4. de Freitas, R., Viana, P.: A graph calculus for proving intuitionistic relation algebraic equations. In: Cox, P., Plimmer, B., Rodgers, P. (eds.) Diagrams 2012. LNCS (LNAI), vol. 7352, pp. 324–326. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31223-6_40

---

[2] More precisely, these logics are typically *defined* by the lack of these rules.

5. Gentzen, G.: Untersuchungen über das logische Schließen I. Math. Z. **39**, 176–210 (1935)
6. Howse, J., Stapleton, G., Taylor, J.: Spider diagrams. LMS J. Comput. Math. **8**, 145–194 (2005)
7. Linker, S., Burton, J., Jamnik, M.: Tactical diagrammatic reasoning. In: UITP 2016. EPTCS, vol. 239, pp. 29–42. Open Publishing Association (2017)
8. Mineshima, K., Okada, M., Takemura, R.: Two types of diagrammatic inference systems: natural deduction style and resolution style. In: Goel, A.K., Jamnik, M., Narayanan, N.H. (eds.) Diagrams 2010. LNCS (LNAI), vol. 6170, pp. 99–114. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14600-8_12
9. Negri, S., von Plato, J., Ranta, A.: Structural Proof Theory. Cambridge University Press, Cambridge (2001)
10. Shin, S.J.: The Logical Status of Diagrams. Cambridge University Press, Cambridge (1995)
11. Stapleton, G., Masthoff, J.: Incorporating negation into visual logics: a case study using Euler diagrams. In: VLC 2007, pp. 187–194. Knowledge Systems Institute (2007)
12. Urbas, M., Jamnik, M., Stapleton, G.: Speedith: a reasoner for spider diagrams. J. Log. Lang. Inform. **24**(4), 487–540 (2015)

# Empirical Studies and Cognition

# Metro Map Colour-Coding: Effect on Usability in Route Tracing

Peter B. Lloyd[1](✉), Peter Rodgers[1], and Maxwell J. Roberts[2]

[1] School of Computing, University of Kent, Darwin Road,
Canterbury CT2 7NZ, UK
{pbl,pjr}@kent.ac.uk
[2] Department of Psychology, University of Essex,
Wivenhoe Park, Colchester CO4 3SQ, UK
mjr@essex.ac.uk

**Abstract.** Does the choice of colour-coding scheme affect the usability of metro maps, as measured by the accuracy and speed of navigation? Using colour to differentiate lines or services in maps of metro rail networks has been a common practice around the world for many decades. Broadly speaking, there are two basic schemes: 'route colouring', in which each end-to-end route has a distinct colour, and 'trunk colouring', in which each major trunk has a distinct colour, and the individual routes inherit the colour of the main trunk that they run along. A third, intermediate scheme is 'shaded colouring', in which each trunk has a distinct colour, and each route has a distinct shade of that colour. In this study, 285 volunteers in the US were randomised to these three colour-coding schemes and performed seventeen navigational tasks. Each task involved tracing a route in the New York City subway map. Overall, we found that route colouring was significantly more accurate than the trunk- and shaded-colouring schemes. A planned subset analysis, however, revealed major differences between specific navigational hazards: route colouring performed better only against certain navigational hazards; trunk colouring performed best against one hazard; and other hazards showed no effect of colour coding. Route colouring was significantly faster only in one subset.

**Keywords:** Metro maps · Colour coding · Navigational hazard
New York City subway · Vignelli · Usability testing

## 1 Introduction

We report on the first part of a project to study the effect of colour coding on the usability of metro maps. A map designer's choice of which colour-coding scheme to use—in particular, colour-coding by individual route versus trunk—is a long-standing point of controversy, but it has not previously been subject to systematic empirical study.

The subway system of New York City has in the past switched between route-colouring and trunk-colouring, and currently has an official online map that shows individual routes but is coloured by trunk. It is therefore an ideal vehicle for investigating the effect of colour-coding scheme in a realistic system. The New York

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 411–428, 2018.
https://doi.org/10.1007/978-3-319-91376-6_38

City subway has one of the most complex service patterns in the world, and is currently undergoing re-evaluation and development of its information delivery, so it is especially suited to the present study. Although this study specifically addresses a metro map of New York City, we believe the results have wider application to metro maps in other cities.

### 1.1 Basic Concepts and Nomenclature

Metro maps exhibit a great variety, as shown by Ovenden (2015). Nomenclature is also varied, and so we will explicitly define the terms that we will be using, as follows.

**Metro maps** are diagrams of urban rail networks, in which the nodes represent stations, edges represent services, and the layout, colouring, and graphical symbolism are designed to assist the passenger in understanding and navigating the network, in order to use it to travel around the city. Thus the map must show both (a) the topological connectivity between the stations as effected by the physical tracks and the passenger services that run along those tracks, and (b) the approximate relationship of the network to the geographic layout of the city.

A **line** comprises one, two, three, or four tracks running together: the tracks that make up a line are usually laid side-by-side, but sometimes are vertically one above the other. In a **two-track line**, the tracks go in opposite directions, 'up' and 'down'; in a **four-track** line, two tracks go up (one express and one local) and two go down (again, express and local); in a **three-track** line, the third track runs express and alternates between up and down in the morning and evening rush hours; in a **single-track** shuttle, the train alternates direction. A **local** train stops at all stations; an **express** train stops at some and skips others. A **route** is a named individual train service, which normally has a single terminal at each end. A **trunk** is a bundle of routes carried on the same line. A **path** is a series of one or more segments of routes, connected by **transfers**. A **journey** is a project by a rider to travel between origin and destination stations. A **run** is a single-route segment of a path between transfers. More informative metro maps show the routes as well as the lines, and a fully informative metro map will show the variation of route service patterns over time of day and day of week.

**Route tracing** is the mental activity of following a single route as it wends its way through a metro map, whereas **journey planning** is the mental activity of navigating a journey based on the map; it comprises 'path construction', that is, assembling a path from one or more route segments, and 'path selection', that is, choosing the best one of several alternative paths. A **navigational hazard** is an identifiable local feature of map that increases the likelihood of a user's making a mistake in these activities.

The map to be examined in this study is an official diagram of the subway network of New York City, used on the MTA (Metropolitan Transportation Authority) **Weekender** web site (MTA 2017). The map was designed by Yoshiki Waterhouse (formerly of Vignelli Associates, now at Waterhouse Cifuentes Design) based on an original 1972 design by Massimo Vignelli, and maintained under Chuck Gordanier, MTA. This diagram was chosen because it is a large and complex map whose structure is suited to experimental changes of colour scheme. Although primarily an online map, the same diagram is increasingly being used in print. This study compares three differently coloured experimental variants of this Weekender map. Figures 1, 2 and 3 show excerpts from the three variant maps used in this study. A **route-coloured** map (Fig. 1) shows each route in a distinctive colour, although this may involve re-using colours, as there may be more routes in total than easily distinguishable colours. **Trunk-coloured** maps (Fig. 2) show each trunk in a distinctive colour: all the routes that run along the same trunk share the same colour. Finally, a **shaded-colour** map (Fig. 3) shows each trunk in a distinctive colour, and each route in a different shade of that colour. Note that a line with no trunk-and-branch structure might still have route colours distinct from trunk colours because the line might carry routes with different express and local stopping patterns. **Hybrid** maps have a mix of trunk and route colouring: this is seen in complex maps, e.g. Cologne (KVB 2017). A hybrid map may be referred to as trunk- or route-coloured if one style predominates.

A map can be **route-drawn** (routes drawn as separate lines) or **trunk-drawn** (the trunk is drawn as one line, with branches). A route-drawn map can be route-coloured, trunk-coloured, or shade-coloured; but a trunk-drawn map must be trunk-coloured. The London Underground map is an example of a predominantly trunk-drawn diagram, while the Weekender is a route-drawn, trunk-coloured map.

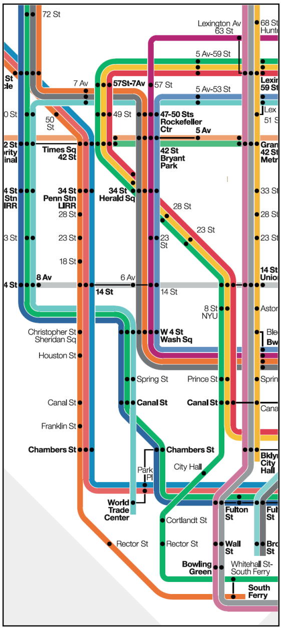


**Fig. 1.** Route colours (Color figure online)

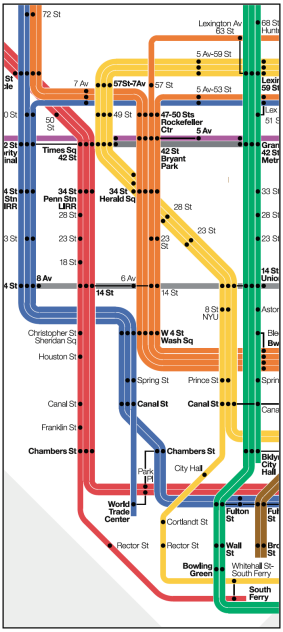


**Fig. 2.** Trunk colours (Color figure online)

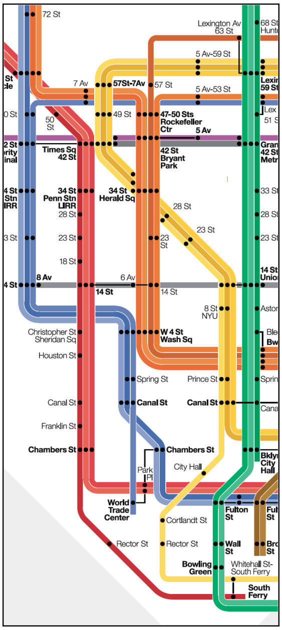


**Fig. 3.** Shaded colours (Color figure online)

**Use of Colour Coding.** Colour coding of metro services started in 1907, when underground railways had already been in use for half a century (Roberts 2012). Different cities adopted the concept differently (Lloyd 2012; Ovenden et al. 2008). For example, in London, Paris, and New York, lines were first coloured by operating company, with trunk colouring used within each line. When those lines were subsumed under municipal control (in 1933, 1930, and 1940 respectively) the maps initially retained that historic colour-coding method. In London, that colour scheme mostly continues today; in Paris, it was replaced by route colouring in 1934; in New York, it was replaced by a route-drawn, route-coloured map in 1967; and by a trunk-drawn, trunk-coloured map in 1979 (Lloyd 2012). In 2011, the MTA introduced the Weekender, a route-drawn map that preserved the 1979 trunk colouring.

### 1.2 Usability

**Historical Usability Studies.** Metro maps have traditionally been created in an artisanal manner without recourse to psychological theory or empirical studies of usability. In recent years, though, a notion of the evidence-based design of metro maps has emerged, combining psychological models with empirical studies of the relationships between observable characteristics of the map and objective measures of the map's usability.

An early study of the usability of a metro map was carried out by the marketing firm Barrington (1966), commissioned by the New York City Transit Authority. Their survey comprised riders' opinions about a prototype design of the map, plus a small set of navigation tests. In the following decade, publicity around the Vignelli map prompted new studies in 1973 by Bronzaft et al. (1976), who selected test journeys with the intention of showing that Vignelli's new schematic map had navigational hazards. She did not intend the study to establish any principles of map design, or to produce an objective evaluation of the Vignelli map in relation to some alternative design.

Although these studies did not provide specific insights into the hazards or design of metro maps, we may at least glean a corroboration of anecdotal reports that enough users find it hard to use metro maps (or, at least, the New York City subway map) to warrant further investigation to characterize and quantify the problem.

Bronzaft and Schachter (1978), Garland et al. (1979) found that colour-coded maps were more usable, and more strongly preferred, than non-colour-coded maps, but they did not address what effect the choice of particular colour-coding schemes might have.

**Modern Usability Studies.** The systematic investigation of metro maps following scientific principles was instigated comparatively recently. Guo (2011) found that the London Underground map layout led to sub-optimal journeys. Roberts (2014a), Roberts and Vaeng (2016) report on the usability effects of large-scale features of the map layout in the London Underground, the London Docklands Light Railway (DLR), the Paris Metro, and the Berlin U-Bahn. The work described in this paper extends those studies by addressing colour coding and local navigational hazards.

Bronzaft in the 1970s used a mix of *in vivo* travel and pencil-and-paper navigation, and modern studies have mostly used pen-and-paper tasks (Roberts et al. 2016) with some use of touch-sensitive screens (Roberts and Rose 2016). Guo et al. (2017) extended the methodology by using the automated presentation of map tests remotely over the internet via Mechanical Turk, a technique that we have adopted in the present study.

The main results yielded by these contemporary usability studies (summarised by Roberts 2014a) are: that objective and subjective evaluations of usability are uncorrelated; that layout affects journey planning; that Beck's octilinear layout is not a 'gold standard' for map layout in all cities, and that instead a framework of more general principles for map layout can be identified, involving the simplicity, coherence, and harmony of line trajectories, and the balance and topographicity of the map layout. This is in stark contrast to the formerly prevailing Beck-centric conventional wisdom (e.g. Field and Cartwright 2014). Which aspects of a map have the most impact on objective usability is an open question for empirical research. Factors likely to influence usability are: overall layout; colour coding; symbolism and layout of transfer stations; symbolism of non-transfer stations; junction layout; and positioning and typography of labels.

**Theory.** Until recently, cognitive theory focused on topographic rather than diagrammatic maps (Robinson 1952; MacEachren 1995; Montello 2002). Stemming from the original work of Miller (1956), the notion of cognitive load is a key to understanding metro maps. For example, Gallotti et al. (2016) have applied cognitive theory to quantify the information overload in map navigation in megacities. In the present project, our model is: (a) tracing a route is essentially a perceptual task, while (b) journey planning is cognitive and involves constructing and comparing alternative paths. Roberts (2014b) outlines two features of map reading: *cognitive load* and *attention capture*. Individuals can handle limited information ('cognitive load'). If the input exceeds that limit, accuracy and speed decline abruptly. Conversely, the salient elements of a task must capture and retain attention. To reduce cognitive load, and heighten attention on salient features, the map should minimise irrelevant noise and be organised to allow the user to find and grasp relevant information. The use of separate route colours is expected to enhance the perceptual task of route tracing but degrade journey planning because their added cognitive load would outweigh the perceptual advantage of distinct colours. Here we test just the first, perceptual part of this theory, that is, route tracing.

**Prediction.** The perceptual task of visually tracking something is easier if that thing can be picked out with distinctive attributes, in this case distinctive colour; and the countervailing disadvantage, namely that route-colouring adds visual clutter to the map as a whole, does not come into play because the subject is not performing a cognitive task. When we isolate route tracing from the more general problem of journey planning, we expect it to be more accurate and faster with route-colouring than with trunk-colouring map; and shaded-colouring to be intermediate. Note: We consider only route-drawn maps and do not address the effect of route-drawn v trunk-drawn layouts.

**Motivation.** The study's impetus comes from a 2014 workshop of wayfinding professionals, which suggested that route-coloured maps would have higher usability. Its relevance comes from metros that switch between colour schemes (e.g. Frankfurt, New York City), plus those that use shaded-colour schemes (Kick 2017; KVB 2017). So there is theoretic and practical interest in the question: what effect does colour scheme have on metro map usability—especially its accuracy, as this has more real-life impact?

## 2 Method

**Map Sourcing.** The MTA provided a vector image file of the subway map of New York City, current at the start of this study (September 2015), and we modified the colours in Adobe Illustrator for this project. We wanted our results to apply to both screen and print media, but the latter has a narrower range of colour tones than screens, so we adjusted the trunk colours of the Weekender map (see below). Our three schemes were:

- **Route-coloured map** used RGB colours approximating the Pantone colours of the 1972 Vignelli map. Some routes did not exist in 1972 and have no 'correct' Unimark colours and so were assigned subjectively reasonable colours (Fig. 1).
- **Trunk-coloured map** closely matched the colours of the Weekender, toned down to the printable range, and adjusted to maintain visibility of station markers (Fig. 2).
- **Shade-coloured map** was inspired by Jabbour's KickMap, but the colours were toned down to the printable range, and adjusted to maintain tonal separation. In each trunk, light and dark shades were alternated for greater contrast (Fig. 3).

We omitted route labels from the map to prevent the route-tracing task becoming trivial. And we omitted background details that were not pertinent to the navigation task (wheelchair access and bus routes) to reduce possible sources of extraneous variance.

We defined test journeys #1 to #17 by selecting seventeen pairs of stations, so that each journey would present one main navigational hazard (Table 1) characterised thus:

- **Route slips:** *flipping* ('*Slip F*') where two or more routes switch places (Fig. 4); *joining* ('*Slip J'*) where two or more routes converge into one trunk on a shared segment of line (three instances are shown in the orange routes in Fig. 5), *underpassing* ('*Slip U*'), where one trunk passes underneath another (Fig. 6), the *twin hazard* ('*Slip T*') of flipping on an underpass (Fig. 7), and *parallel running* ('*Slip P*') where two or more routes have a long run together (Fig. 8).

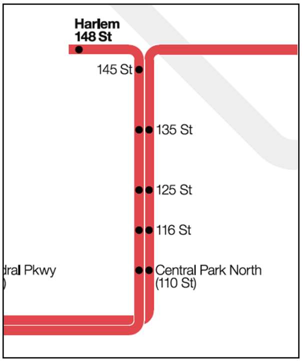


**Fig. 4.** Slip F, #12

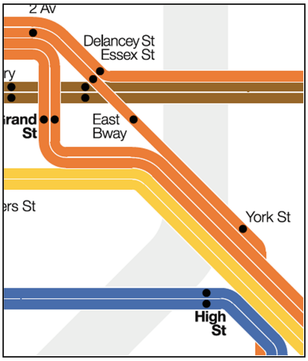


**Fig. 5.** Slip J, #8/#9

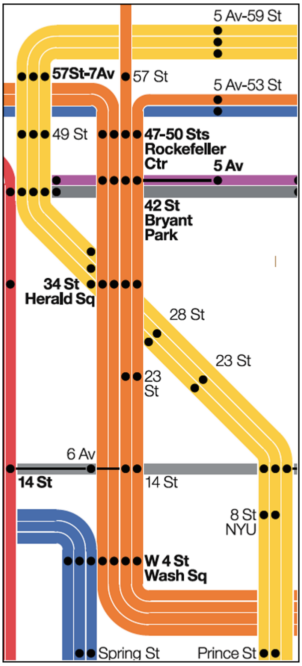


**Fig. 8.** Slip P, #7

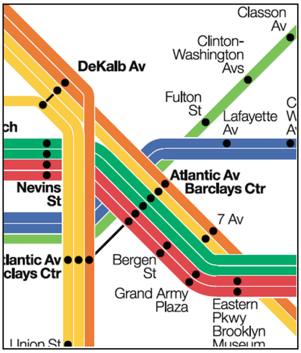


**Fig. 6.** Slip U, #10/#11

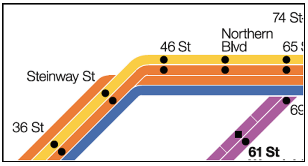


**Fig. 7.** Slip T, #14/#15

- **Branch jumps:** these comprise *route jumps* (‘*Jump R*’) where the rider ‘jumps’ across a branched route, in this case Beach 90 St to Beach 44 St (Fig. 9) and *trunk jumps* (‘*Jump T*’) where the rider jumps across a branched trunk, 138 St-Grand Concourse to 3 Av-138 St in this case (Fig. 10).

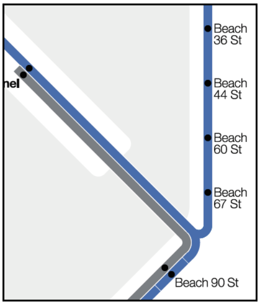


**Fig. 9.** Jump R, #2

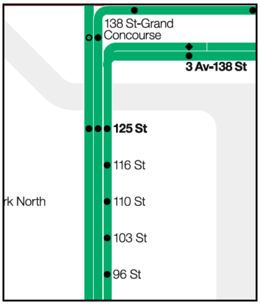


**Fig. 10.** Jump T, #4

- **Mistransfers:** The only mistransfer considered was the *phantom transfer* (*'Trans P'*), where the user believes a transfer is needed when it is not, in this case when travelling from 168 St to 157 St (Fig. 11).

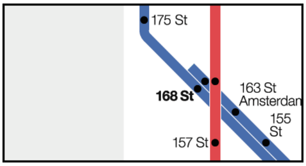


**Fig. 11.** Trans P, #17

Four attention exercises were added, which were trivial journeys to check that the user was paying attention and not just clicking at random, #18 to #21.

**Table 1.** Routes.

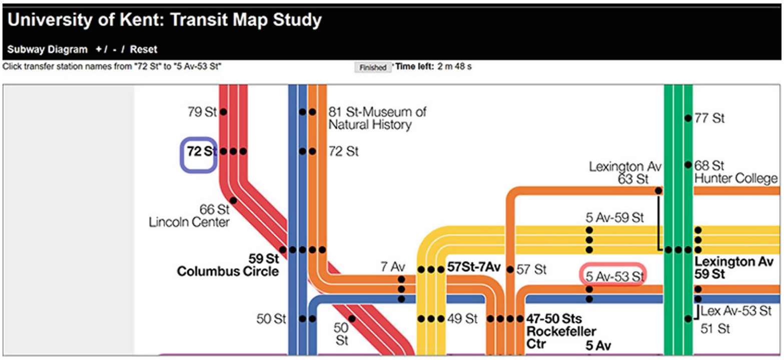
| # | From | To | Hazard |
|---|---|---|---|
| 1 | 104 St (A) | Aqueduct \| North Conduit Ave (A) | Jump R |
| 2 | Beach 90 St (A) | Beach \| 44 St (A) | ″ |
| 3 | Nereid Av (5) | Eastchester \| Dyre Av (5) | ″ |
| 4 | 138 St-Grand \| Concourse (4,5) | 3 Av \| 138 St (6) | Jump T |
| 5 | 72 St (B,C) | 5 Av-53 St (M,E) | ″ |
| 6 | 175 St (A) | 155 St (B,D) | ″ |
| 7 | Prince St (R,N) | Queensboro \| Plaza (N,Q,7) | Slip P |
| 8 | Coney Island \| Stillwell Av (D,N,F,Q) | Wood\|haven Blvd (R,M) | Slip J |
| 9 | Coney Island \| Stillwell Av (D,N,F,Q) | Forest Hills\|-71 Av (R.F.M,E) | ″ |
| 10 | Spring St (C,E) | Lafayette\|Av (C) | Slip U |
| 11 | Spring St (C,E) | Nostrand Av (A,C) | ″ |
| 12 | New Lots Av (3) | Harlem \| 148 St (3) | Slip F |
| 13 | New Lots Av (3) | 149 St-Grand Concourse (2,4,5) | ″ |
| 14 | Jamaica-179 St (F) | Court Sq \| 23 St (M,E) | Slip T |
| 15 | Jamaica-179 St (F) | Roosevelt \| Island (F) | ″ |
| 16 | City Hall (R) | Queensboro \| Plaza (N,Q,7) | ″ |
| 17 | 168 St (A,C,1) | 157 St (1) | Trans P |
| 18 | Euclid AV (A,C) | Grant Av (A) | Attention |
| 19 | Kings Hwy (N) | Coney Island \| Stillwell Av (D,N,F,Q) | ″ |
| 20 | Fulton \| St (G) | Clark St (3) | ″ |
| 21 | Kingsbridge \| Rd (4) | Woodlawn (4) | ″ |

**Recruitment.** Subjects were recruited from across the USA through the Amazon *Mechanical Turk* crowdsource service (Amazon 2017) and paid $7.50 for up to an hour's work. Such crowdsourcing tools have been established as a platform for conducting behavioural science experiments (Crump et al. 2013; Paolacci et al. 2010). Whilst laboratories provide a controlled space for experimenters, they are typically restricted in participant numbers and demographics (Gadiraju et al. 2017). Crowdsourcing provides access to a large number of diverse participants (Mason and Suri 2012). Issues can arise: first, a lack of control of participants' attention to the task, who might be clicking randomly to complete the task as quickly as possible; second, the lack of a means of interacting with the participant for training. We mitigated these, in the first case through attention-checking questions that could be performed without much cognitive effort, and in the second case through careful design of the study software to include a sizeable automated training phase (which was developed in the supervised pre-pilot study).

**Study Software.** We presented the maps as Scalable Vector Graphic (SVG) images in JavaScript within Mechanical Turk. As each subject began, the software randomly allocated a colour scheme (trunk, route, or shaded). It then presented four predefined training trials, which had the same appearance as the test trials, but the subject was told whether his/her answer was correct. For each test trial, the software randomly selected one of the twenty-one origin-destination pairs, and displayed them in a zoomed rectangle of the map, with blue and red outlines around the start and end station names. The words "Start" and "Finish" were displayed (in blue and red respectively) in a large font (40 pixels) to draw attention to these stations, these two words fading over ten seconds.

Above the map was the question, "Can you travel from station <$S_1$> to station <$S_2$> without changing trains?" (e.g. Fig. 12), followed by radio buttons labelled "Yes" and "No". When the user clicked on one them, a third button appeared, labelled "Finish", which the user clicked to close that trial and go to the next. There were three clickable symbols immediately above this, "+" and "−" for zoom in and out, and "Reset", and users could pan the map by clicking to grab it—all reproducing the interface on the Weekender site. In a laboratory pre-pilot study, we found that users had difficulty using these controls, so zooming by thumbwheel was added. An online pilot study using Mechanical Turk showed no further changes were needed, so we moved to the full study.

Software and data: https://www.cs.kent.ac.uk/projects/metromap/.


University of Kent: Transit Map Study

Click transfer station names from "72 St" to "5 Av-53 St"

Time left: 2 m 48 s

Fig. 12. Sample screenshot (for test journey from 72 St to 5 Av–53 St) (Color figure online)

Each subject was allowed three minutes per trial: any timeouts were logged as a ‘failure by timeout’ (and counted as a non-success in the statistics), and the subject moved to the next trial. A total of 10 trials were timed out (4 trunk, 4 route, 2 shaded).

**Exclusions.** Neither the training trials nor the attention exercises were included in the statistics presented here. We excluded all data from any subject who failed in two of the four attention exercises, or who failed to complete the full set of tasks, or who completed the tasks but failed to upload the data to Mechanical Turk. Of 305 subjects who were randomised, 20 were excluded for those reasons.

## 3 Results

Navigation accuracy when following a route in the New York City diagrammatic subway map is significantly affected by the colour-coding scheme. Table 2 and Fig. 13 shows the analysis of 4,845 trials by 285 subjects (98 trunk, 93 route, 94 shade colouring). Each subject’s accuracy is scored as the percentage of correct answers in 17 trials.

**Table 2.** Effect of colour coding on accuracy & speed of navigation.

| Mean score (%), S ± *SE* | | | | Mean time (seconds), T ± *SE* | | | |
|---|---|---|---|---|---|---|---|
| Trunk | Route | Shaded | | Trunk | Route | Shaded | |
| 66.8 ± *1.4* | 71.9 ± *1.2* | 65.7 ± *1.3* | p < 0.01,<br>F(2,282) = 5.7 | 26.8 ± *1.0* | 24.5 ± *1.1* | 27.3 ± *1.4* | NS,<br>F(2,282) = 1.5 |

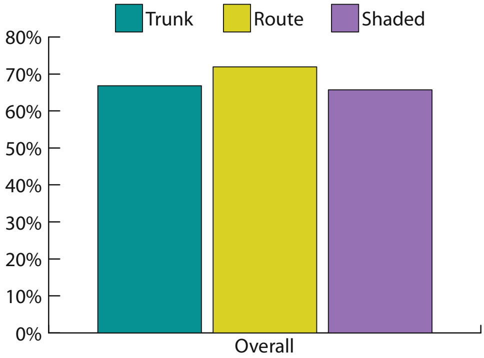


**Fig. 13.** Mean score

Route colouring was the most accurate, while trunk colouring was only slightly more accurate than shaded. A one-factor, three-level analysis of variance showed a statistically significant effect of colour scheme on accuracy ($p < 0.01$), although only one pair-wise comparison (route v shaded) was significant ($p < 0.05$). Route colouring was also the fastest, but the effect of colour coding on time was not statistically significant.

The main navigational hazard presented in each trial affected the accuracy of navigation. We examined two classes of hazard: slippage between routes, and jumping across branches (Table 3 and Fig. 14). In slippage, route colouring's advantage was more pronounced, but in branch jumping, the effect was inverted and trunk performed best.

**Table 3.** Subset analysis by general class of navigational hazard.

| | Mean score (%), S ± *SE* | | | | Mean time (seconds), T ± *SE* | | | |
|---|---|---|---|---|---|---|---|---|
| | Trunk | Route | Shaded | | Trunk | Route | Shaded | |
| Slip | 68.4 ± *1.5* | 81.1 ± *1.5* | 71.3 ± *1.6* | $p < 0.01$,<br>F(2,282) = 15.8 | 31 ± *1.2* | 29.4 ± *1.3* | 31.8 ± *1.6* | NS,<br>F(2,282) = 0.8 |
| Jump | 69.0 ± *2.3* | 60.0 ± *1.9* | 61.9 ± *2.3* | $p < 0.01$,<br>F(2,282) = 4.8 | 21.0 ± *1.0* | 18.0 ± *1.0* | 21.8 ± *1.4* | NS,<br>F(2,282) = 3.0 |

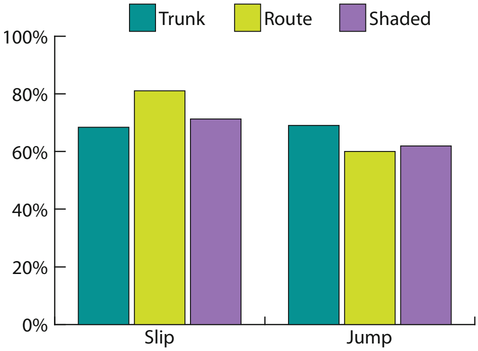


**Fig. 14.** Mean score, by class of navigational hazard

In both subsets Slip and Branch, a three-level analysis of variance showed a significant effect of colour coding (both $p < 0.01$). In the Slip subset, pair-wise analysis of variance showed that route colouring is significantly more accurate than either trunk or shaded (both $p < 0.01$) but trunk and shaded are not significantly different from each other. In the Jump subset, pair-wise analysis of variance shows only that trunk colouring is significantly more accurate than route ($p < 0.05$).

Analysis of eight specific types of navigational hazard indicated a strong dependence of the colour-coding effect on the main navigational hazard that is presented in a trial: three out of five slippage hazards showed large and significant advantages in route colouring, while the other two showed no significant dependence on colour coding (Table 4 and Fig. 15). The two branch-jumping hazards showed a significant effect, but with opposite directions–one favouring trunk colouring, the other route. Phantom transfers showed worse-than-chance error rates, with a significant advantage in route colouring.

In most subsets, route colouring yielded the fastest navigation, but only in one ('Jump T') did this attain statistical significance.

Where the main navigational hazard is route slippage due to flipping (Slip F), joining (Slip J) and the twin hazard of flipping on an underpass (Slip T), route colouring shows a strong and statistically significant advantage in accuracy of navigation. Trunk colouring is worst, and shaded colouring has intermediate accuracy. In each of these three subsets, three-level analysis of variance shows a significant effect of colour scheme ($p < 0.01$); and pair-wise analyses show that route colouring is significantly more accurate than trunk ($p < 0.01$) or shaded ($p < 0.05$) but trunk and shaded are not significantly different from each other.

**Table 4.** Subset analysis by specific type of navigational hazard

| | Mean score (%), S ± *SE* | | | | Mean time (seconds), T ± *SE* | | | |
|---|---|---|---|---|---|---|---|---|
| | Trunk | Route | Shaded | | Trunk | Route | Shaded | |
| Slip F | 70.9 ± *3.1* | 89.2 ± *2.4* | 75.0 ± *3.3* | $p < 0.01$,<br>F(2,282) = 9.9 | 38.1 ± *2.1* | 38.9 ± *2.4* | 39.7 ± *2.9* | NS,<br>F(2,282) = 0.1 |
| Slip J | 52.6 ± *2.3* | 68.8 ± *2.6* | 56.4 ± *2.7* | $p < 0.01$,<br>F(2,282) = 10.1 | 53.6 ± *3.1* | 43.6 ± *3.0* | 52.3 ± *3.7* | NS,<br>F(2,282) = 2.6 |
| Slip T | 65.3 ± *3.0* | 85.7 ± *2.1* | 71.3 ± *2.2* | $p < 0.01$,<br>F(2,282) = 15.8 | 22.5 ± *1.1* | 20.9 ± *1.2* | 22.4 ± *1.3* | NS,<br>F(2,282) = 0.5 |
| Slip P | 79.6 ± *4.1* | 75.3 ± *4.5* | 80.9 ± *4.1* | NS,<br>F(2,282) = 0.5 | 23.1 ± *1.5* | 26.0 ± *2.2* | 26.4 ± *2.1* | NS,<br>F(2,282) = 0.8 |
| Slip U | 80.6 ± *3.5* | 81.2 ± *3.5* | 77.7 ± *3.7* | NS,<br>F(2,282) = 0.27 | 20.3 ± *1.0* | 19.8 ± *1.1* | 20.2 ± *1.3* | NS,<br>F(2,282) = 0.0 |
| Jump R | 46.3 ± *3.7* | 23.7 ± *3.6* | 35.8 ± *3.5* | $p < 0.01$,<br>F(2,282) = 9.25 | 21.4 ± *1.2* | 18.6 ± *1.1* | 21.6 ± *1.6* | NS,<br>F(2,282) = 1.5 |
| Jump T | 91.8 ± *1.9* | 96.4 ± *1.5* | 77.7 ± *2.2* | $p < 0.01$,<br>F(2,282) = 4.9 | 20.6 ± *1.1* | 17.4 ± *1.1* | 22.1 ± *1.5* | $p < 0.05$,<br>F(2,282) = 3.5 |
| Trans P | 37.8 ± *4.9* | 51.6 ± *5.2* | 33.0 ± *4.8* | $p < 0.05$,<br>F(2,282) = 3.68 | 14.2 ± *1.0* | 15.4 ± *1.1* | 15.8 ± *1.5* | NS,<br>F(2,282) = 0.5 |

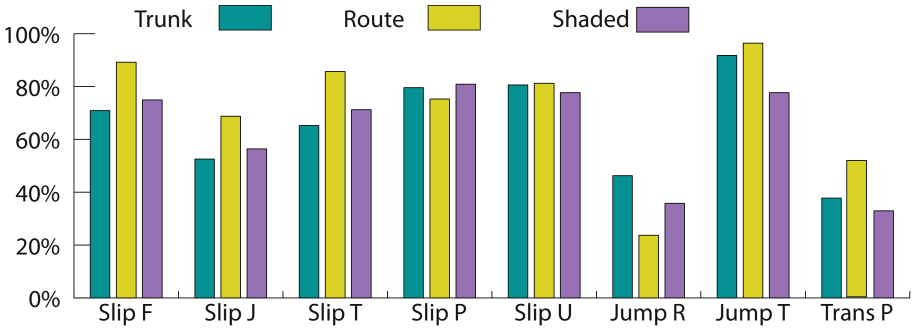


**Fig. 15.** Mean score, by individual type of navigational hazard

In the trials where the main navigational hazard is route slippage due to parallel running (Slip P) or underpassing (Slip U), neither the accuracy nor the speed of navigation exhibited any significant dependence on the colour coding.

Branch jumping exhibits quite different behaviour from route slippage. In the trials where the main navigational hazard is jumping across a route branch (Jump R), both the accuracy and speed of navigation are strongly dependent on colour scheme, but—unlike in the slippage hazard—the trunk colour scheme is most accurate and route colouring worst. Note that these results are all worse than chance: in effect, the shaded- and route-coloured maps are positively misleading, while the trunk-coloured map is almost as good as tossing a coin. In a three-level analysis of variance, colour coding has a significant effect on accuracy ($p < 0.01$), and pair-wise analysis of variance shows that trunk is significantly better than route colouring ($p < 0.01$). The other pair-wise comparisons are not statistically significant.

In the trials where the main navigational hazard is jumping across a trunk branch (Jump T), the error rate is low, but the pattern is the same as in the slippage hazards: route colouring is the most accurate. Three-level analysis of variance shows that colour coding has a significant effect on accuracy ($p < 0.01$); pair-wise analysis of variance shows that route colouring is significantly better than shaded ($p < 0.05$). The other pair-wise comparisons are not statistically significant.

In both of these Jump subsets, route colouring is the fastest. Three-level analysis of variance shows a significant effect of colour coding only for Jump T. Pair-wise analyses of variance are not statistically significant.

In the trials where the only hazard is a phantom transfer, we have a situation where, like Jump R above, the map is positively misleading with success rates substantially below chance, and there is a significant dependence on colour-coding scheme. In contrast with Jump R, however, in Trans P, route colouring is most accurate. Three-level analysis of variance showed a significant effect of colour coding ($p < 0.05$); but the pair-wise analyses of variance are not significant. Speed of navigation shows no clear pattern in this subset, and analysis of variance yields nothing statistically significant.

# 4 Discussion

## 4.1 Navigational Hazards

**Route Slippage.** The results confirm the central plank of our theoretical platform: the navigational task of tracing a line as it weaves through the subway map is more accurately performed when done with a map that employs route colouring. It was expected that shaded colouring would yield an intermediate accuracy between those of route colouring and trunk colouring. In fact, in these tasks there is no real difference in accuracy between the trunk and shaded colouring.

Route colouring's advantage is mainly manifested in two recognisable navigational hazards: flipping, where two routes swap their positions within a trunk (referred to as 'Slip F'); and joining, where routes join a trunk ('Slip J'). Two other hazards that were expected to be comparable were also analysed: underpassing, where one trunk passes under another trunk ('Slip U'); and parallel running, where a trunk runs for a long distance with multiple bends but without any of the other hazards ('Slip P'). In all four hazards, we expected the perceptual task of tracing one route in a trunk of identically coloured routes would be less accurate than doing so when each trunk is distinctly coloured. In fact, the hazards Slip U and Slip P showed no differential effect of colour coding. A fifth subset was analysed, whose principal risk was the twin hazard of flipping immediately next to an underpass: as expected, this yielded a higher error rate than Slip F or Slip U alone, and exhibited a significant effect of the colour coding, but this appears to be due entirely to the flipping, not the underpassing.

It appears that when the parallel routes closely follow each other (as in Slip P and Slip U), it is easier for the eye to follow them, and route colouring offers no advantage; but when the routes behave differently (as in Slip F and J), they are harder to follow as they split the user's attention, and then route colouring offers an advantage.

**Branch Jumping.** We expected that colour coding might affect the reading of branches, and that subjects might mistakenly believe they can 'jump' across junctions without changing trains. We therefore included six trials to investigate 'branch jumping': three to test for jumping across trunk junctions, where a route splits off from a trunk ('Jump T'), and three for jumping across route junctions, where a route splits into two limbs ('Jump R'). A clear result was obtained: Jump R had a high error rate, and route colours performed worst; Jump T had a low error rate, and route colours performed best. It is curious that Jump R is affected by the colour scheme, despite the fact that the junction itself looks the same whichever scheme is used. Only the context provided by the rest of the map changes. It is hypothesised that in a route-coloured map, the user is confused by the inconsistent use of the colour coding in the branch, with two undifferentiated routes given the same colour, and the confusion disrupts the thinking and causes errors.

**Mistransfers.** Besides the test trials, four trivial trials were included to provide a check on the subject's attention. In the laboratory pre-pilot study, however, one of these tasks, the phantom transfer ('Trans P'), had surprisingly high error rates and showed a strong effect of colour coding. This was therefore added as the seventeenth test trial in the online pilot study and the main study; and a replacement attention trial added. The cause of the errors seems to be a misreading of the Vignelli transfer symbolism at 168 St station, at the junction of the 7 Ave Broadway Line (formerly IRT) and the Washington Heights line (BMT), where the three station dots are not aligned. As this navigational task does not involve tracking routes over any long distances or through any perceptual hazards, it is surprising that accuracy in this task was strongly affected by colour coding. Its high error rates mean that, in a trunk- or shaded-coloured map, this junction is positively misleading. (With route colours, the passenger has an even chance of reading it correctly.) We speculate that trunk and shaded colouring increases the perceived separation of station markers on the two lines, as the two Broadway routes are linked by the same colour, while the Washington Heights route is differentiated by its colour.

**Speed.** Our original hypothesis was that route colouring would enhance the speed of navigation as well as its accuracy. The results neither prove nor disprove the hypothesis, but they do suggest the alternative hypothesis that the tracking component of following a route takes the same time whether it is done correctly or incorrectly. The only subset where colour coding had a significant effect on speed was Jump T, which involves no tracking but the more cognitive task of understanding a junction. It is expected that a correlation between accuracy and speed will emerge in the more cognitive task of journey planning, which we will address in the second part of this project.

### 4.2 Implications for Design

The original impetus for this research was the practical design question of whether to draw a metro map with route or trunk colour coding. It was hypothesised that route colouring would make it easier to carry out the navigational task of tracing a route, but that the visual clutter it creates would make it harder to carry out complex journey planning. It was an open question whether one or the other of those two factors would predominate and therefore whether it is possible to recommend one or the other colour scheme. The present study examines only the route-tracing task; a further study of the effect of colour coding on journey planning is in preparation.

We have found that colour coding does not have a consistent effect across the board. It depends on navigational hazard. Route colouring offers its greatest advantage in tracing routes against the hazards of flipping and joining. If a metro map has a lot of those hazards, but is not so complex that it requires difficult journey planning, then we would expect route colouring to yield a more usable map.

A more general outcome is the identification of distinct navigational hazards, which a designer may address in mapmaking, and software should avoid in automated design.

- Flipping causes the user to lose track of a route, especially in a trunk-coloured map. Change the layout to avoid flips or develop less hazardous flips.
- Joining routes within a trunk causes the user to lose track of a route: make the separation clearer by changing the layout or the colouring.
- If routes are drawn separately in the map, then splitting a route creates a high risk of branch jumping. Never split routes. This argues against hybrid maps.
- If transfers are shown with Vignelli-style dot proximity then always keep the dots aligned, and try to avoid drawing such stations at junctions.

### 4.3 Implications for Automated Design

Work on the automated design of metro maps (e.g. Stott and Rodgers 2004; Wolff 2007) has centred on algorithmic solutions to the mathematical problem of laying out a metro map subject to multiple criteria of what constitutes of a good layout. The results of the present study indicate that besides the overall layout, the colour coding and specific local navigational hazards need to be addressed. Algorithmic methods for automatically finding and fixing local navigational hazards could have an important role to play.

### 4.4 Implications for Further Research

Inspection of the data from this study reveals that users often make poor choices, either journeys that are longer than necessary, or that involve more transfers than necessary, or that involve attempts to execute non-existent routes or transfers. Work by Roberts has led to a framework for the 'macro design' of a map layout. The present work points to a need for a comparable investigation of 'micro design' to establish the effect of colour, transfer layout and symbolism, and junction layout on usability.

## 5 Conclusions

There is a strong interaction between the colour-coding scheme and usability in the diagrammatic New York City subway map, as measured by accuracy of navigation when carrying out route-tracing tasks. This can be understood in terms of specific navigational hazards. In a broad class of hazards (Slip F, Slip J, Slip T, Jump T, and Trans P), the route colouring has a large, statistically significant advantage in accuracy. In one type of hazard (Jump R), the pattern is reversed, and trunk-colouring outperforms route-colouring; in two cases (Slip P, Slip U), colour coding has no significant effect.

In terms of the psychological model: We conclude that route-coloured maps become more accurate than trunk- and shaded-coloured maps when routes that belong to the same trunk behave differently by branching and switching places, and thereby divide the user's attention, as opposed to simply running in parallel. We also conclude that route-coloured maps become less accurate if the colour coding is applied inconsistently (for example, route splitting), thereby confusing the user's thinking.

These results indicate that local navigational hazards can play a major role in the overall usability of a metro map. And that the choice of a colour-coding scheme that would yield the most usable map in real-life usage must depend on the comparative frequency of different kinds of hazard in the map, and the comparative frequency of different navigations (route tracing *versus* journey planning) undertaken by passengers.

**Acknowledgement.** The MTA very generously allowed the use of the Weekender map in this study, and provided a digital copy of the map.

## References

Amazon.com, Inc.: Mechanical Turk. https://www.mturk.com/mturk/. Accessed 28 Nov 2017

Barrington and Company: A Survey of Prospective User Reaction to the New Subway Map, 31 May 1966, report submitted to the NYCTA. Collection of Raleigh D'Adamo (1966)

Bronzaft, A.L., Dobrow, S.B., O'Hanlon, T.J.: Spatial orientation in a subway system. Environ. Behav. **8**, 575–594 (1976)

Bronzaft, A.L., Schachter, J.: Memo to MTA Subway Map Committee, 3 March 1978. New York Transit Museum (1978)

Crump, M.J.C., McDonnell, J.V., Gureckis, T.M.: Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS ONE **8**(3), e57410 (2013). https://doi.org/10.1371/journal.pone.0057410

Gadiraju, U., Möller, S., Nöllenburg, M., Saupe, D., Egger-Lampl, S., Archambault, D., Fisher, B.: Crowdsourcing versus the laboratory: towards human-centered experiments using the crowd. In: Archambault, D., Purchase, H., Hoßfeld, T. (eds.) Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments. LNCS, vol. 10264, pp. 6–26. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66435-4_2

Field, K., Cartwright, W.: Becksploitation: the over-use of a cartographic icon. Cartogr. J. **51**(4), 343–359 (2014). https://doi.org/10.1179/0008704114Z.000000000150

Gallotti, R., Porter, M.A., Barthelemy, M.: Lost in transportation: information measures and cognitive limits in multilayer navigation. Sci. Adv. **2**, e1500445 (2016). https://doi.org/10.1126/sciadv.1500445

Garland, H.C., Haynes, J.J., Grubb, G.C.: Transit map color coding and street detail: effects on trip planning performance. Environ. Behav. **11**(2), 162–184 (1979)

Guo, Z.: Mind the map! The impact of transit maps on path choice in public transit. Transp. Res. Part A **45**, 625–639 (2011). https://doi.org/10.1016/j.tra.2011.04.001

Guo, Z., Zhao, J., Whong, C., Wyman, L.: Redesigning subway map to mitigate bottleneck congestion: an experiment in Washington DC using Mechanical Turk. Transp. Res. Part A **106**, 158–169 (2017). https://doi.org/10.1016/j.tra.2017.09.017

Kick Design: KickMap. http://kickmap.com/. Accessed 29 Nov 2017

KVB (Kölner Verkehrs-Betriebe AG): Bahnen in Köln (2017). http://www.kvb-koeln.de/german/fahrplan/linienplan.html. Accessed 29 Nov 2017

Lloyd, P.B.: Vignelli: Transit Maps. RIT Press, Rochester (2012)

MacEachren, A.M.: How Maps Work: Representation, Visualization, and Design. Guilford Press, New York (1995)

Mason, W., Suri, S.: Conducting behavioral research on Amazon's Mechanical Turk. Behav. Res. Methods **44**(1), 1–23 (2012). https://doi.org/10.3758/s13428-011-0124-6

Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. **63**(2), 81–97 (1956)

Montello, D.M.: Cognitive map-design research in the twentieth century: theoretical and empirical approaches. Cartogr. Geogr. Inf. Sci. **29**(3), 283–304 (2002)

MTA (Metropolitan Transportation Authority, New York): The Weekender. http://web.mta.info/weekender.html. Accessed 29 Nov 2017

Ovenden, M.: Transit Maps of the World. Penguin, New York (2015)

Ovenden, M., Lloyd, P.B., Pepinster, J.: Paris Métro Style. Capital Transport, Harrow (2008)

Paolacci, G., Chandler, J., Ipeirotis, P.G.: Running experiments on Amazon Mechanical Turk. Judgm. Decis. Making **5**(5), 411–419 (2010)

Roberts, M.J.: Underground Maps Unravelled. Self-published, Wivenhoe, UK (2012)

Roberts, M.J., Newton, E.J., Canals, M.: Radi(c)al departures: comparing conventional octolinear versus concentric circles schematic maps for the Berlin U-Bahn/S-Bahn networks using objective and subjective measures of effectiveness. Inf. Des. J. **22**(2), 92–115 (2016). https://doi.org/10.1075/idj.22.2.04rob

Roberts, M.J.: Schematic maps in the laboratory. In: First International Schematic Mapping Workshop University of Essex, 2–3 April 2014 (2014a). http://www.tubemapcentral.com/articles/Roberts_empirical_SMW2014_2014.pdf. Accessed 29 Nov 2017

Roberts, M.J.: What's your theory of effective schematic map design? ibid (2014b). http://repository.essex.ac.uk/11368/1/Roberts-Theoretical.pdf. Accessed 29 Nov 2017

Roberts, M.J., Rose, D.: Map-induced journey-planning biases for a simple network: a docklands light railway study. Transp. Res. Part A **94**, 446–460 (2016). https://doi.org/10.1016/j.ijhcs.2016.06.003

Robinson, A.H.: The Look of Maps. University of Wisconsin Press, Madison (1952)

Stott, J.M., Rodgers, P.: Metro map layout using multicriteria optimization. In: Proceedings of 8th International Conference on Information Visualisation (IV 2004), 14–16 July 2004, pp. 355–362. IEEE (2004). https://doi.org/10.1109/iv.2004.1320168, https://kar.kent.ac.uk/14133. Accessed 29 Nov 2017

Wolff, A.: Drawing subway maps: a survey. Inform.—Forschung Entwickl. **22**, 23–44 (2007). https://doi.org/10.1007/s00450-007-0036-y

# The Effect of Graphical Format and Instruction on the Interpretation of Three-Variable Bar and Line Graphs

Nadia Ali and David Peebles(✉)

Department of Psychology, University of Huddersfield,
Queensgate, Huddersfield HD1 3DH, UK
`{n.ali,d.peebles}@hud.ac.uk`

**Abstract.** We present a study that investigates how graph format and training can affect undergraduate psychology students' ability to interpret three-variable bar and line graphs. A pre and post-test design was employed to assess 76 students' conceptual understanding of three-variable graphs prior to and after a training intervention. The study revealed that significant differences in interpretation are produced by graph format prior to training; bar graph users outperform line graph users. Training also resulted in a statistically significant improvement in interpretation of both graph formats with effect sizes confirming the intervention resulted in substantial learning gains in graph interpretation. This resulted in bar graph users outperforming line graph users pre and post training making it the superior format even when training has occurred. The effect of graph format and training differed depending on task demands. Based on the results of this experiment, it is argued that undergraduate students' interpretations of such three-variable data are more accurate when using the bar form. Findings also demonstrate how a brief tutorial can result in large gains in graph comprehension scores. We provide a test which can be used to assess students understanding of three-variable graphs and the tutorial developed for the study for educators to use.

## 1 Introduction

Analysing and interpreting quantitative data is a key skill taught in all scientific undergraduate degree courses because the ability to work with data is a fundamental activity in the sciences [1]. Although different skills are important for students to master, one vital skill in the development of scientific inquiry is the ability to work with quantitative data [2] The expectation that people should be able to read and interpret basic data has progressed to an expectation that individuals can actively work with the data and manipulate information depending on the nature of scientific inquiry [3]. Active interpretation of data requires skills which allow a reader to make inferences from given data, find trends, criticise data and use data to support and evaluate claims. Therefore proficiency in

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 429–440, 2018.
https://doi.org/10.1007/978-3-319-91376-6_39

data literacy in today's information age is a necessary pre-requisite to scientific inquiry skills [1].

The expanding utilisation of visual presentation of information in science, the media and regular daily life depends on the presumption that charts and graphs are straightforward to the viewer, due to the human capability of recognising a pattern and inferring the quantitative relationship being depicted [4]. However, reading scientific graphs requires more than encoding of pattern [5] and when progressing beyond the interpretation of simple pattern relationships [6,7], the utility of the representation will depend on an interaction between the individual's graphical literacy, the graph format used and whether the format supports the task the reader is required to engage in [8–13].

Although many experienced graph users take their abilities for granted, the knowledge and skills required are far from trivial and require considerable training and practice to be mastered [3,9,14]. A large body of research investigating graph reading ability has revealed that novice students misinterpret scientific graphs and that most errors can be traced to a deficit in perceptual and conceptual understanding of how the visualisation represents information [4,5,15]. For example, a consistent and ubiquitous finding in the physics education literature [16] is that students exhibit misconceptions such as interpreting graphs literally (as pictures) and 'slope-height' confusion where students incorrectly assume a greater slope implies a higher value.

Similar findings concerning consistent misconceptions have been found with graphs representing data from experimental designs depicting the effect of a one or more independent variables on a dependant variable. These type of experimental designs are very prevalent in psychology; a subject where students are required to learn the fundamentals of experimental design and statistical analysis of one or more independent variable on a dependant variable. These designs are known as "factorial research designs".

## 2 Factorial Research Designs

Factorial research designs are widely used in all branches of the natural and social sciences as well as in engineering, business and medical research. The efficiency and power of such designs to reveal the effects and interactions of multiple independent variables (IVs) or factors on a dependent variable (DV) has made them an invaluable research tool and, as a consequence, the teaching of such designs, their statistical analysis and interpretation lies at the core of all natural and social science curricula.

The simplest form of factorial design is the two-way factorial design, containing two factors, each with two levels, and one DV (for example the differences in wellbeing (DV) between men and women ($IV_1$) as a function of high and low exercise regimes ($IV_2$)). Statistical analysis of these designs most often results in a $2 \times 2$ matrix of mean values of the DV corresponding to the pairwise combination of the two levels of each IV. Interpreting the results of even these simplest of designs accurately and thoroughly is often not straightforward however, but requires a significant amount of conceptual understanding—for example the concepts of 'simple', 'main' and 'interaction' effects [9]. Like most statistical analyses, interpretation can be eased considerably by representing the data in diagrammatic form [17–19]. Data from two-way factorial designs are most often presented as either three-variable line or bar graphs—variously called 'interaction' or 'ANOVA' graphs. Examples of such bar and line graphs (taken from the experiments reported here) are shown in Fig. 1. Consistent with findings in the domain of physics, research investigating these three-variable graphs reveals a systematic bias in interpretation centred on two-variables with a deficiency of interpretation concerning the third variable.

Although the graphs displayed here are relatively simple (depicting the relationship between three variables) research has consistently revealed that the majority of graph viewers will struggle to interpret them accurately when the information is depicted in line graph format [5,15,20]. Despite relatively minor differences between the two types of graphs experimental studies have revealed that line graph users were significantly more likely to misinterpret or be unable to interpret the data represented than bar graph users [15]. In previous research we hypothesised that these observed differences result from a combination of two factors: (a) the inadequate knowledge structures and procedures of some novice users and (b) Gestalt principles of perceptual organisation [21] that made data points and their relationships more visually salient in the bar graphs, thereby making their interpretation, particularly by novices, much more detailed and accurate [15,20].

Specifically, the visual salience of the lines in the line graphs drew attention to them and readers could associate the line pattern to the legend via a colour matching process. However, the line connecting the data points made the identification and interpretation of the specific data points relating to the variable plotted on the x axis more difficult. These errors were less likely to be found in bar graph interpretations because each data point is represented by a unique, readily identifiable bar. To test this notion we designed a novel line graph design [15] to offset the bias present in traditional line graphs. Data points were coloured and matched to their corresponding variables by placing a colour patch next to the associated variable on the x axis. Consistent with the analysis once this novel colour match line graph shared similar anchoring principles as the bar graph format performance was equivalent for both graph formats [15].

One potential implication of these findings is that three-variable data of this type may be more effectively taught to undergraduate students in the form of bar graphs than with the more traditional line graphs (or the modified graph design of the colour match graph). Based on previous findings it would appear that bar graphs are superior to line graphs when presenting statistical information to a student population. A possible longer term implication may be that this recommendation is more generally applicable to other forms of data.

## 3 The Effect of Training on Statistical Reasoning

There still remains however an important question regarding the robustness of this effect and whether training can produce any discernible benefits. Although research has been conducted demonstrating that design modification [15,22] and cognitive scaffolding [23] can boost accuracy and quality of interpretation, no studies which we are aware of have investigated the effect of direct training on comprehension of three-variable bar and line graphs and how training may potentially interact with graph format.

Additionally, although working with visual representations of data is considered to be an essential skill in scientific reasoning and there is an increasing demand in the literature for these skills to be taught [3] we could not find any tutorial guiding students on how to interpret three-variable Cartesian co-ordinate graphs or graphical tests which assessed comprehension of these types of graphs. Graph comprehension is a complex task [14] so it is often the case that novices will not benefit from the purpose of the visualisation, or worse the representation will increase misconceptions and erroneous interpretations of data [20]. Dreyfus and Eisenberg (1990, p. 33) argue that: “Reading a diagram is a learned skill; it doesn’t just happen by itself. To this point in time, graph reading and thinking visually have been taken to be serendipitous outcomes of the curriculum. But these skills are too important to be left to chance” [14].

A systematic review of the literature [3] concluded that graph interpretation and construction had to explicitly be taught in order for graduate students to develop scientific inquiry skills in data handling and interpretation. The level of skill needed to appropriately interpret data from graphs depends on the task demands on the user. These task demands have traditionally been classified as elementary, intermediate and advanced [1,3] in the literature and increasing sophistication of skills is associated with higher educational achievement. Elementary reasoning is the simplest and requires the user to simply read the data by locating specific information from the graph. For example, a point reading question for the graph in Fig. 1a would be “How much $CO_2$ do Quebec plants which are chilled uptake?” The graph user is then expected to read the information from the graph and accurately locate that the $CO_2$ uptake is 50 units.

Intermediate reasoning involves identifying the relationship between variables and trends being depicted in the graph. For example an intermediate reasoning question for the graph in Fig. 1a would be “Describe how the treatment affects each plant type?”. The user is then expected to describe the relationship between variables, a step up from reading information off such as point reading. An example of an intermediate interpretation is: “In the case of chilled treatment both plants take up the same amount of $CO_2$ but for non-chilled treatment Mississippi takes up a lot more $CO_2$ than Quebec”. Advanced reasoning involves extrapolating from the data such as generalising to a population, making a prediction based on the trend or a comparison of trends and variable groupings [1,3]. In factorial research designs advanced reasoning involves identifying main effects of each independent variable (e.g., for the graph in Fig. 1b, “Overall fasting results in a much higher glucose uptake than not fasting”) and if there is an interaction effect present. An example of an interpretation of an interaction effect in the graph in Fig. 1b is "When you are fasting relaxation training slightly increases glucose uptake. When you are not fasting relaxation training slightly decreases glucose uptake. Therefore the effect of relaxation training on glucose uptake reverses depending on whether fasting occurred or not".

In previous research we have demonstrated that novices may be able to provide interpretations of the graphical pattern but do not have the knowledge structures to be able to explicitly identify main and interaction effects (indeed it is only in an expert sample such advanced reasoning occurs [9,20]). In addition, we also found that novice students struggle with elementary interpretation if they cannot relate the pattern to the variables the pattern represents [9,20]. It is crucial therefore that the rules of graphical representations are taught or even basic reasoning may be difficult for a non-expert audience of graph users. To address this need for training this paper describes an experimental intervention where students were taught how to interpret these graphs depicting results of factorial research designs. In order to assess graph comprehension prior to training and after training, pre- and post-tests designed to measure graph reading ability were also developed. Both measures are described in more depth below.

It may be the case that a high rate of error in graph interpretation emerges in the absence of appropriate and explicit instruction. If so, the conclusion to be drawn would be that explicit and rigorous teaching of line graph interpretation is essential in statistics to prevent it being hampered by the potentially confusing features of the format. Alternatively, it is possible that the visual salience of the lines in line graphs is so high that its effect on interpretation is still found after explicit training has occurred. If this is the case, then it may be wise to conclude that such data would be best taught and communicated in bar graph form. The key questions this study aims to address are:

1. Is one particular graph format more appropriate than another for students in Further and Higher Education?
2. What effect (if any) does a training intervention have on students' ability to reason with graphical information?
3. How does the effect of graph format and training differ depending on task demands?
4. Is there an interaction effect between graph format and training?

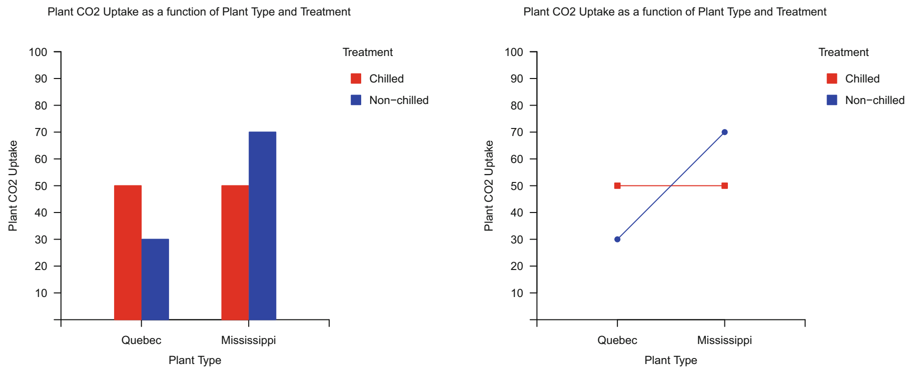
## 4 Method

### 4.1 Participants

Participants were 80 foundation level undergraduate psychology students at the University of Huddersfield with 40 participants in each graph condition. There were 36 participants who completed both the pre and post-test in the line graph condition and 40 participants who completed both in the bar graph condition making the overall sample size 76.

## 4.2 Materials

Two tests were constructed using two data sets that produced two pairs of bar and line graphs. These graphs were informationally equivalent in that no information can be inferred from one that cannot be inferred from the other and each can be constructed from the information in the other [19]. In addition, all features were identical between the two graph formats apart from the pattern in the centre connecting the data points. The graphs for the session 1 test are shown in Figs. 1a and 1b while those for session 2 are shown in Figs. 1c and 1d[1].



(a) Bar graph – pre-training (b) Line graph – pre-training

(c) Bar graph – post-training (d) Line graph – post-training

**Fig. 1.** Factorial graph stimuli used in the pre-training and post-training tests

---

[1] The pre-test is available at http://peebles.sdfeu.org/heapn/q1b.pdf and the post-test is available at http://peebles.sdfeu.org/heapn/q2l.pdf.

The variables in the graphs used for the test were chosen so that no prior knowledge of the domain or relationships would influence interpretation. The values of the conditions were devised to present and test the fundamental items of knowledge required to produce an appropriate knowledge structure or schema for each graph and to produce patterns that would test the various hypotheses under investigation. The questions in both questionnaires were essentially identical, with only minor changes in wording to account for the different graph formats. The questions were devised to examine students' knowledge of relevant concepts at an elementary, intermediate and advanced level. Elementary questions (questions 1–4) probed for knowledge of independent and dependent variables, correct identification of causal relationships and point reading questions. The maximum score for these set of questions was 9.

These questions were followed by questions 5–12 which required intermediate reasoning. Specifically the questions required a simultaneous consideration of the two independent variables to establish the effect on the dependant variable, e.g., mean values, minimum and maximum values etc.). The maximum score for these set of questions was 16. Advanced reasoning involved questions which probed knowledge of main effects, an interaction effect and ability to be able to consider every combination of the levels of each IV on the DV (questions 13–15). The maximum score for these set of questions was 6. Both sessions 1 and 2 graphs showed a possible main effect of the independent variable plotted on the x-axis and an interaction effect.

Similar to the test, a tutorial was also developed as an instructional intervention to teach students how to interpret these three-variable graphs (one for line graphs and the other for bar graphs). The information in both tutorials was essentially identical, with only minor changes to account for the different graph formats. The tutorial mirrored the test of graphicacy and covered basic to advanced skills in graph reading and statistical information extraction. Therefore the tutorial begins with elementary instruction, such as where the independent and dependant variables are plotted, how to associate pattern to variables, etc. then progresses onto intermediate reasoning (how to simultaneously consider the effect of two IV's on a dependant variable, how to transform data to provide mean scores etc.) and advanced instruction which focussed on how to establish whether a main effect and an interaction effect is present[2].

For the purpose of the experiment the tutorial was delivered as a 25 min presentation (in a lecture theatre, during a first year cognitive psychology class) by the two authors who practised delivery prior to the experiment and read off standardised scripts to ensure consistency between conditions.

### 4.3 Design and Procedure

The study consisted of three elements; a pre-test, an instructional intervention and a post-test 2 weeks later. A mixed design was employed to assess the effect of graph format and instruction on graph comprehension. An independent group

---

[2] The tutorial is available at: http://peebles.sdfeu.org/heapn/heapn-tutorial.swf.

design was employed to assess the effect of graph format, with different participants being given bar or line graphs. The test was a repeated measures design where participants completed both the pre-test and the post-test.

The test was in either bar or line graph form and was randomly distributed to students which resulted in random allocation to each experimental condition (bar or line graph). They were then immediately given a 25 min lecture on graph interpretation at the same time by each respective author in separate lecture rooms. We chose this setting (lecture hall where students attend teaching sessions) to increase the validity of the learning material and the learning environment. However, employing elements of a field study meant that counterbalancing of graphs was not possible in the pre and post-test design.

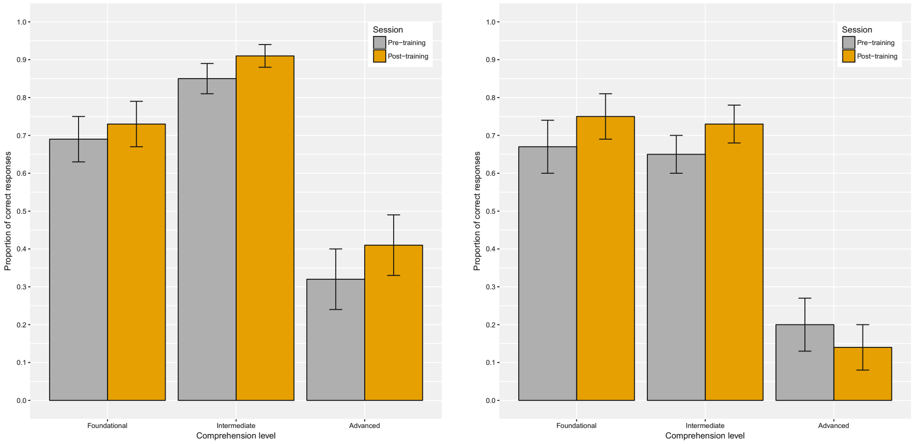
The lecture was simply a presentation delivered of the tutorial produced. Students were informed that they needed to remember whether they were assessed using the bar or line graph format. A period of 14 days separated the two tests after which the students completed the second test, again in class. The students informed the authors the graph condition they had been allocated to, and the authors did a check of pre and post-tests to ensure they were completed by the same person.

### 4.4 Scoring

Each question was scored as correct or incorrect by the author. Where questions had multiple response options (e.g., name the independent variable(s)) negative marking was employed to control for guessing and to prevent inflation in scores. The maximum overall score which could be obtained on the test is a score of 31.

## 5 Results

The results are discussed in terms of effect size as well as statistical significance which allows for a meaningful consideration of the results in an educational context. Descriptive analysis (Fig. 2) reveals bar graph users outperform line graph users before and after training, with the one exception being the post-test scores assessing foundation reasoning whereby performance is very similar in both groups. Therefore this format is superior to the line graph format for depicting three-variable data sets. Training itself improves performance although the benefit differs depending on task demand. Therefore the improved effect of training interacts with task demands, improvement in intermediate reasoning is more pronounced than improvement in foundation or advanced reasoning. Variance is similar in both conditions apart from when intermediate reasoning is being assessed, in which case variance is much higher in the line graph condition compared to the bar graph condition. Therefore, in a student sample performance is better when the bar graph format is used and there is more consistency in performance when intermediate reasoning is required if this format is used (Fig. 2).



(a) Bar graph condition

(b) Line graph condition

**Fig. 2.** Proportion of correct responses for each comprehension level and testing session, bar and line graph conditions. Error bars indicate 99% confidence intervals.

A mixed method ANOVA determined that test scores differed significantly as a function of graph format, $F(1,74) = 14.613, p < .001, \eta_p^2 = .165$, and this difference is large—bar graph users scored on average 17% higher than line graph users. Training also resulted in statistically significant improvement, $F(1,74) = 15.230, p < .001, \eta_p^2 = .171$, and the effect size is also large, indicating a substantial benefit from the educational intervention. There was no interaction effect between graph format and training, $F(1,74) = 0.033, p = .855, \eta_p^2 = 0.0$.

To establish whether improvement from training differed as a function of task demands and graph format, paired samples t-tests were conducted using a Bonferroni corrected alpha of 0.017. These revealed that in the line graph condition there was a statistically significant improvement from pre to post-test when foundation reasoning is assessed ($p = .009$) and when intermediate reasoning is assessed ($p = .01$) but not when advanced reasoning is required ($p = .134$).

In the bar graph condition training did not produce significant gains in foundation reasoning ($p = .177$) but produced significant improvement in intermediate reasoning ($p = .007$). The adjustment of the alpha level results in training not producing a statistically significant improvement in advanced reasoning ($p = .022$).

## 6 Discussion

The experiment presented in this article provides insight into how graph format, task demands and training in graph interpretation affects the comprehension of three-variable bar and line graphs. The experiment revealed three key findings which have important implications for graphical display design and educational recommendations for which graph format to use when presenting data from statistical analyses.

First, informationally equivalent bar and line graphs are not computationally equivalent for students in higher education. Bar graph users are more likely to correctly interpret information than line graph users prior to any training. This effect is stable, irrespective of task demands. This finding has now been replicated using different methods to assess comprehension including verbal protocols, written responses and question answer tasks of the type used in this study [15,20]. The effect size is large and indicates that 17% of variation in performance on the test can be accounted for by graph format, consistent with previous findings that novice students perform substantially better when the representation they work with is in bar form. The increased benefit of training for bar graph users means that this format still surpasses the line format post-training (Fig. 1).

Secondly a brief training intervention designed to improve graph comprehension results in a marked improvement when the results are considered in the context of a one off tutorial in graph interpretation lasting 25 min. The effect size is large, indicating that the educational intervention resulted in significant improvement, especially in the line graph condition. Training results in improvements in foundation reasoning when data is presented in the line graph form but not in bar graph form. This is consistent with previous findings that novice students struggle with elementary reasoning when data is presented in the line graph form but not the bar form. Specifically novice users struggle to correctly associate the pattern in the centre to variables plotted on the axes [15,20,24]. Once this simple matching process has been taught through the tutorial a significant improvement emerges in foundation and intermediate reasoning when using the line graph form.

Advanced reasoning requires a long time to develop [9]; novices are unlikely to have the knowledge structures to assist them in identifying main effects and interaction effects [15,20]. However, the improved performance in the bar graph condition extends to advanced reasoning indicating some benefit from the tutorial. Therefore the component of the tutorial providing instruction on advanced reasoning would require additional study, although study can be tailored around the individual student's test score using our pre and post tests. The video can also be treated as a hyperlink so components of the tutorial can be targeted for re-study such as components involving advanced reasoning.

## 7 Summary and Recommendations

Bar and line graphs are some of the most commonly used graphical formats for presenting data from some of the most commonly used statistical tests in the social sciences [25]. Our research findings demonstrate that degree level students perform better when using bar graphs than when using line graphs. The effect of graph format is substantial even without training. It is reasonable to assume that further training would result in instruction accounting for a greater variance in performance.

Training results in improvements in reasoning with both graph formats. Therefore, the recommendations are clear: students should use bar graphs when interacting with visual displays depicting quantitative data and a brief tutorial can improve reasoning with this format to a considerable extent. Higher education institutions can use our tutorial to provide such training whilst teaching statistics as part of the psychology degree program. The training can be tailored around the sophistication of skills. For example, a student may find one off instruction is sufficient for them to be able to manage elementary reasoning, but repeated practise would be necessary to develop advanced skills such as identifying holistic trends such as main effects and interaction effects. We also provide a test of graphicacy for HEI's to assess students' reasoning with graphs presenting results of experimental designs.

**Acknowledgements.** This research was supported by a grant from the UK Higher Education Academy Psychology Network.

## References

1. Friel, S.N., Curcio, F.R., Bright, G.W.: Making sense of graphs: critical factors influencing comprehension and instructional implications. J. Res. Math. Educ. **32**, 124–158 (2001)
2. National Council of Teachers of Mathematics: Principles and Standards for School Mathematics, vol. 1. National Council of Teachers (2000)
3. Glazer, N.: Challenges with graph interpretation: a review of the literature. Stud. Sci. Educ. **47**(2), 183–210 (2011)
4. Pinker, S.: A theory of graph comprehension. In: Freedle, R. (ed.) Artificial Intelligence and the Future of Testing, pp. 73–126. Lawrence Erlbaum Associates, Hillsdale (1990)
5. Shah, P., Carpenter, P.A.: Conceptual limitations in comprehending line graphs. J. Exp. Psychol. Gen. **124**, 43–62 (1995)
6. Cleveland, W.S., McGill, R.: An experiment in graphical perception. Int. J. Man-Mach. Stud. **25**(5), 491–500 (1986)
7. Zacks, J., Tversky, B.: Bars and lines: a study of graphic communication. Mem. Cogn. **27**(6), 1073–1079 (1999)
8. Okan, Y., Garcia-Retamero, R., Cokely, E.T., Maldonado, A.: Improving risk understanding across ability levels: encouraging active processing with dynamic icon arrays. J. Exp. Psychol.: Appl. **21**(2), 178–194 (2015)
9. Peebles, D., Ali, N.: Expert interpretation of bar and line graphs: the role of graphicacy in reducing the effect of graph format. Front. Psychol. **6**, 1673 (2015)
10. Peebles, D., Cheng, P.C.H.: Modeling the effect of task and graphical representation on response latency in a graph reading task. Hum. Factors **45**, 28–45 (2003)
11. Carswell, C.M., Wickens, C.D.: The perceptual interaction of graphical attributes: configurality, stimulus homogeneity, and object interaction. Percept. Psychophys. **47**, 157–168 (1990)
12. Carswell, C.M., Wickens, C.D.: Mixing and matching lower-level codes for object displays: evidence for two sources of proximity compatibility. Hum. Factors **38**(1), 1–22 (1996)

13. Peebles, D.: The effect of emergent features on judgments of quantity in configural and seperable displays. J. Exp. Psychol.: Appl. **14**(2), 85–100 (2008)
14. Dreyfus, T., Eisenberg, T.: On difficulties with diagrams: theoretical issues. In: Booker, G., Cobb, P., de Mendicuti, T.N. (eds.) Proceedings of the 14th Annual Conference of the International Group for the Psychology of Mathematics Education, vol. 1, pp. 27–36 (1990)
15. Ali, N., Peebles, D.: The effect of gestalt laws of perceptual organisation on the comprehension of three-variable bar and line graphs. Hum. Factors **15**(1), 183–203 (2013)
16. Heckler, A.F.: The ubiquitous patterns of incorrect answers to science questions: the role of automatic, bottom-up processes. In: Mestre, J.P., Ross, B.H. (eds.) Psychology of Learning and Motivation: Cognition in Education, vol. 55, pp. 227–268. Academic Press, Cambridge (2011)
17. Kosslyn, S.M.: Graph Design for the Eye and Mind. Oxford University Press, New York (2006)
18. Field, A.: Discovering Statistics Using SPSS, 4th edn. Sage, London (2009)
19. Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. Cogn. Sci. **11**, 65–100 (1987)
20. Peebles, D., Ali, N.: Differences in comprehensibility between three-variable bar and line graphs. In: Taatgen, N., van Rijn, H., Nerbonne, J., Schoemaker, L. (eds.) Proceedings of the Thirty-first Annual Conference of the Cognitive Science Society, Mahwah, NJ, pp. 2938–2943. Cognitive Science Society (2009)
21. Wertheimer, M.: Laws of organization in perceptual forms. In: Ellis, W.D. (ed.) A Source Book of Gestalt Psychology. Routledge & Kegan Paul, London (1938)
22. Shah, P., Mayer, R.E., Hegarty, M.: Graphs as aids to knowledge construction: signaling techniques for guiding the process of graph comprehension. J. Educ. Psychol. **91**, 690–702 (1999)
23. Mautone, P.D., Mayer, R.E.: Cognitive aids for guiding graph comprehension. J. Educ. Psychol. **99**(3), 640–652 (2007)
24. Ali, N., Peebles, D.: The different effects of thinking aloud and writing on graph comprehension. In: Carlson, L., Holscher, C., Shipley, T. (eds.) Proceedings of the Twentieth Annual Conference of the Cognitive Science Society, Mahwah, NJ. Cognitive Science Society (2011)
25. Beattie, V., Jones, M.J.: The use and abuse of graphs in annual reports: theoretical framework and empirical study. Account. Bus. Res. **22**(88), 291–303 (1992)

# Read It *This* Way: Scaffolding Comprehension for Unconventional Statistical Graphs

Amy Rae Fox and James Hollan

University of California – San Diego, San Diego, CA, USA
amyraefox@ucsd.edu

**Abstract.** How do you make sense of a graph that you have never seen before? In this work, we outline the types of prior knowledge relevant when making sense of an unconventional statistical graph. After observing students reading a deceptively simple graph for time intervals, we designed four instructional scaffolds for evaluation. In a laboratory study, we found that only one scaffold (an interactive image) supported accurate interpretation for most students. Subsequent analysis of differences between two sets of materials revealed that task structure–specifically the extent to which a problem poses a mental impasse–may function as a powerful aid for comprehension. We find that prior knowledge of conventional graph types is extraordinarily difficult to overcome.

**Keywords:** Graph comprehension · Scaffold · Unconventional graph

## 1 Introduction

Imagine when reading a paper you encounter a graph, teeming with information—surely important by virtue of the precious column inches it spans. But as you scan for patterns, willing the author's insight to leap off the page, you find there is something unattainable. Like the writing of a foreign language, you see familiar symbols and structure, yet the rules for assembling these pieces into a meaningful whole are just outside your grasp. How do you make sense of the graphic?

As Larkin and Simon note, "a representation is useful only if one has the productions that can use it," [1, p. 71]. If we lack the ability to draw inferences from a representation, then we may find it largely useless. How is it that we develop such productions for *new* graphical forms, when even familiar systems (like scatter plots and line graphs) can prove challenging to interpret [2]? In this work, we build upon research on reading and graph comprehension to explore how readers make sense of an unconventional statistical graph. After generating hypotheses for instructional scaffolding techniques through observation (Study One), we evaluate their efficacy in the laboratory (Study Two). We find that even with explicit (text or image-based) instructions, the influence of prior knowledge from conventional graph forms is difficult to overcome. Our results suggest that when presenting unconventional graphical forms, effective techniques will direct readers' attention to the salient differences between their expectations and reality, and that designers mustn't take for granted that readers will notice they are dealing with an unconventional form.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 441–457, 2018.
https://doi.org/10.1007/978-3-319-91376-6_40

### 1.1 Cognitive Aids for Graph Comprehension

Owing largely to their importance in STEM education, techniques for supporting graph comprehension have been a focus of research in the learning, cognitive and computer sciences. The most minimal interventions involve *graphical cues*—visual elements that guide attention, akin to gesture and pointing in conversation. Acartürk [3, 4] investigated the influence of point markers, lines and arrows on bar charts and line graphs, finding that different cues can lead readers to interpret a graph as depicting either an event or process. Similarly, Kong and Agrawala [5] proposed the term "graphical overlays" to refer to elements layered onto content to facilitate specific graph-reading tasks. Reviewing a corpus of statistical graphs in popular media they identified five common types of overlays: (1) reference structures (such as gridlines) (2) highlights, (3) redundant encodings (such as data value labels), (4) summary statistics and (5) annotations, each aimed at reducing cognitive load for particular graph-reading tasks.

Turning to more elaborate interventions, Mautone and Mayer [6] investigated techniques from reading comprehension to support meaningful processing of graphs in the college classroom. In a series of experiments, they presented learners with scatterplot and line graphs augmented by *signaling* (animations to reveal components of a graph, adding cues to highlight the relationship of depicted variables), *concrete graphic organizers* (diagrams & photographs of the real-world referents of variables in a graph) and *structural graphic organizers* (diagrams depicting a relationship analogous to the one represented in a graph). They found that the type of cognitive aids provided to learners affected subsequent structural interpretation of the graphs (measured by relational or causal statements).

Importantly however, these studies did not differentiate between prior knowledge of the *domain* and knowledge of the *graphs* [3, 4, 6]. The cognitive aids explored in this literature do not instruct users on *how* to read the graphs – the "rules" for their representational systems. Rather, it is assumed that the reader has familiarity with the type of graph being read. Scatterplots, time series and line graphs all rely on the Cartesian coordinate system, serving as a common graphical framework [7]. We are interested in what happens when presented with a graph that doesn't rely on this framework. Might we need a different type of scaffolding to learn a novel representational system?

### 1.2 Prior Knowledge and Graphical Sensemaking

Modern theories of graph comprehension posit a combination of bottom-up and top-down processing [8]. While the design of a graph is clearly important, so too is the nature of prior knowledge we bring to the task. When making sense of a graph, we draw on at least two sources of prior knowledge: our knowledge of the domain, and of the graphical form [2]. Scarcity from either source will impede comprehension in different ways.

**Limited Prior Knowledge.** If presented with an unfamiliar graph, depicting information in an unfamiliar domain, I will be unable use knowledge of one to bootstrap inferences for the other. Consider a novice physics student reading a Feynman diagram: without the requisite understanding of particle physics, they cannot reverse-engineer the formalisms of the diagram. Without these formalisms, they cannot draw inferences about particle physics.

**Limited Domain Knowledge.** Alternatively, if presented with a familiar graph depicting data in an unfamiliar domain, I might draw on my knowledge of the graph system to learn something new about the content. If I know a straight line represents a linear relationship, I can infer that such a relationship exists between the (unfamiliar) variables in a line graph connected by a straight line [8]. It is this situation we aim to optimize in STEM education. Mautone and Mayer [6] demonstrated that animations, arrows, diagrams and photographs can all help students connect their prior knowledge of graphs to depicted variables, improving their ability to draw inferences about the related scientific processes. Of course, our expectations about how a graph works, if inappropriate, can also lead to systematic errors in interpretation [2].

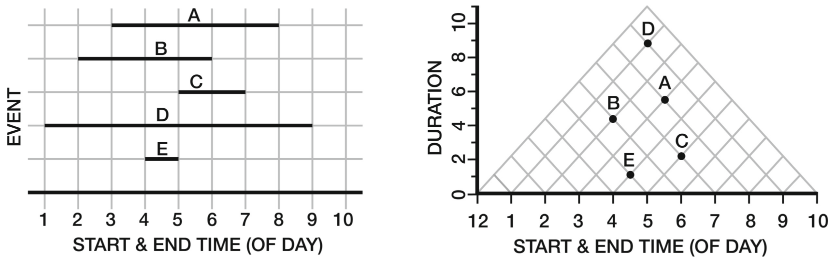
**Limited Graphical Knowledge.** We are interested in the reciprocal case: an unfamiliar representation depicting information in a familiar domain. Importantly, by graphical knowledge we are *not* referring to knowledge of graphs in general–graphical competency–but rather knowledge of the rules governing a *particular* graph form. We reason that existing techniques for scaffolding are insufficient for this case, as the information added to the graphs serve only to strengthen the relationship between the graph-signs and (real-world) referents. This fails to address the learner's scarcity of knowledge for the representational system. If we cannot perform first order readings–such as extracting a data value–we cannot hope to perform second order readings–like inferring relationships between variables.

With sufficient domain knowledge, we expect that learners may be able to reverse-engineer the formalisms governing an unconventional graph. We wish to scaffold this process to support self-directed graph reading. As a first step, we select an obscure graphical form using an unconventional coordinate system so that we might shed light on the graphical framework: the foundation of the graph schema [7].

### 1.3 The Triangular Model of Interval Relations

Several representational systems for reasoning about time intervals have been explored in the literature [9], due largely to their importance in data analysis across the sciences and humanities. We have selected two informationally equivalent [1] types of time interval graphs, each representing the start and end time, duration, and relations between intervals.

In Fig. 1-*left*—the **Linear Model of Temporal Relations** (*hereafter* LM)—intervals are depicted as line segments along a one-dimensional timeline running from left-to-right. The left and right boundary points of a line segment indicate the start and end time, respectively, while the length of the segment indicates its duration. In the LM, the *y*-axis is solely exploited to differentiate between intervals, for example, by use of a label. In this way, the second dimension contains no *metric* information. As a result, intervals can be sorted along the *y*-axis in numerous ways (e.g. by start time, duration, alphabetically by label, etc.). As noted by Qiang et al. [10] this polymorphism prohibits the existence of a standard approach to visual pattern recognition with the LM, making it ill-suited for applications in exploratory data analysis and inspection of extremely large data sets.



A Linear Model (LM) Graph

A Triangular Model (TM) Graph

**Fig. 1.** Informationally-equivalent graphs for intervals of time

Based on work by Kulpa [9] extended by Qiang et al. [10, 11] the **Triangular Model of Temporal Relations** (*hereafter* TM) overcomes this shortcoming by representing intervals as points in 2D metric space (Fig. 1-*right*). Each point represents an interval. In the vertical dimension, the height of the point indicates its duration. The intersection of the point's triangular projections (using diagonally oriented grid lines) onto the *x*-axis indicate the start and end time. In this way, every interval is represented as a unique point in the 2D graph space, and each of its elementary properties are explicitly encoded by the location of the point.

A brief inspection of the TM by even the most experienced graph readers demonstrates its relative obscurity. However, while the non-Cartesian coordinate system is unconventional, the graph depicts information about a domain in which we all share substantial prior knowledge: events in time.

### 1.4 The Present Studies

We are interested in what happens when experienced graph readers (undergraduate STEM majors) attempt to interpret the TM graph. Further, we wish to develop and evaluate a series of instructional scaffolds to support comprehension of the graph by self-directed readers. We start by observing students using the TM graph to solve simple questions about the properties and relations between events, and then elicit suggestions for how to make the graph easier to read (Study One). In Study Two, we evaluate four scaffolds inspired by these observations.

## 2 Study 1: Observing Learning of an Unconventional Graph

What strategies do we employ to make sense of an unconventional graph? In this exploratory study, we observed students solving problems with the Triangular Model (TM) graph (Part A). After a short interview, we challenged students to design instructional aids making the graph easier to read (Part B). From these data we generate hypotheses for how we might scaffold comprehension for novel statistical graphs.

### 2.1 Methods

**Participants.** Twenty-three (70% female) English speakers from the experimental-subject pool at a large American university (M(age) = 20, SD(age) = 1) participated in exchange for course-credit. All students were majors in STEM subjects. Participants were recruited in dyad pairs (9 pairs, n = 18) to encourage a naturalistic think-aloud protocol. In cases where one recruit was absent we conducted the session with the individual (n = 5), altering the procedure only by encouraging them to think-aloud as though explaining their reasoning to a partner. In total, we conducted 14 observation sessions (9 dyads, 5 individuals).

**Materials and Procedure.** The entire procedure ranged from 45–60 min. In Part A: The Graph Reading Task, sixteen multiple choice questions were used to probe the reader's ability to use the graph to reason about the properties of and relations between intervals. For example, a question testing the "duration" property might read: *For how many hours does event [x] last?* Participants were given one sheet of paper containing the questions and a second sheet containing a large TM graph with 15 data points[1]. After delivering instructions, we started the video recording and left the room.

Upon task completion, we conducted a short interview, prompting participants to explain how they would plot a new data point on the graph. If participants misinterpreted the graph, we began a didactic interview, prompting students to ask questions they thought might help them discover the rules of the graph system. We responded by only revealing the information explicitly requested, minimizing the effect our teaching might have on the designs produced in the next task. Once students could plot a new data point, we proceeded to Part B: The Scaffold Design Task. We asked participants to think about what they could do to make the graph easier to read for the next participant and invited them to make marks on the graph.

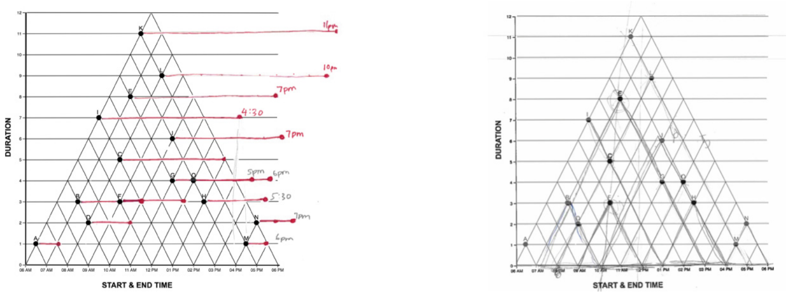
### 2.2 Study One: Results

**Part A. Graph Reading Task.** Participants in only 3 of the 14 sessions correctly interpreted the TM graph (M(score) = 12/16 points, (SD = 1.7), (M(time) = 19 min, SD = 30 s). These participants correctly described the graph's rules in the post-task interview. In the remaining 11 sessions, participants correctly answered only 2.2 questions on average (SD = 2.1), and were unable to correctly plot a point in the interview. Yet in these sessions, participants *did* persist in answering all questions, spending about the same amount of time on the task (M(time) = 21 min, SD = 2 min). Reviewing the artifacts participants generated gives us a window into their interpretations. Looking first at the lowest scoring sessions, we noticed many cases where participants appeared to superimpose the conventional representation for time intervals–the linear model (see Fig. 1-*left*) – atop the triangular graph (Fig. 2-*left*). We dubbed this the "linear interpretation" of the TM, which relies on participants assuming

---

[1] **Note.** All materials, data and computational notebooks for data analysis are available at https://madebyafox.github.io/Scaffolding_Graph_Comprehension

the data points are situated in a Cartesian coordinate system with a single *x* and *y* intercept. They must also infer that a point represents a *moment* in time, rather than an interval, and that the interval is represented by a line segment which they must mentally project (or physically draw) atop the graph. They must also decide which moment along an interval the point represents. In this sense, the "linear interpretation" relies on two kinds of prior knowledge: first of Cartesian coordinates in which a point has a single *x*-intercept, and secondly of conventions for representing intervals as linear extents, rather than points. This interpretation also requires students to ignore—or assign no meaningful referent to—the graph's diagonal gridlines. Once constructed, participants could extract information from the "linear interpretation" following the same procedure one would follow for the conventional linear (LM) graph.



The lowest-scoring session shows an (incorrect) Cartesian interpretation.

The highest-scoring session shows a (correct) triangular interpretation.

**Fig. 2.** Graph artifacts from lowest (*left*) and highest (*right*) scoring sessions.

Alternatively, In Fig. 2-*right* we see the artifact from the highest scoring session. Participants have reinforced the triangular intersections for several points with the *x*-axis. Noticeably, we do not see reinforcement of the intersections with the *y*-axis, presumably because this is a convention of the coordinate system participants did not need assistance to interpret.

**Testing the Linear Interpretation Hypothesis.** From our review of participants' graph markings, as well as the procedure they (initially) described for plotting a new data point, we hypothesized that the 11 low-scoring sessions had formed a "linear interpretation" of the graph. To test this hypothesis, we constructed an alternative answer key. First, we constructed a "linear interpretation" graph by drawing a vertical intersect for each data point to the x-axis and construing this as the *start time*. We then drew horizontal line segments from each point, with a length determined by the *duration* given on the *y*-axis. Using this "linear interpretation" graph, we determined the correct answer for every problem and re-scored each session. Under this alternative answer key, the mean score for the 11 lowest-scoring sessions improved from 2.2 to 8.3 (SD = 2.7 points), while the mean score for the 3 highest-scoring sessions decreased 12.3 to 3.0 (SD = 2.0 points), supporting the hypothesis that low-scoring participants interpreted the graph in accordance with the conventional linear model.

**Part B. Scaffold Design Task.** We evaluated the artifacts produced in response to our prompt to make the graph easier to read, and found evidence of three instructional approaches: adding pictorial intersections (Fig. 3a), providing annotations/examples (Fig. 3b, c) and text instructions (Fig. 3d).

In Fig. 3a (at right) participants have drawn attention to the diagonal gridlines and their dual-intersections with the *x*-axis by darkening and coloring them. These participants explained the most challenging part of the graph was realizing they had to look for *two* intersections with the *x*-axis.

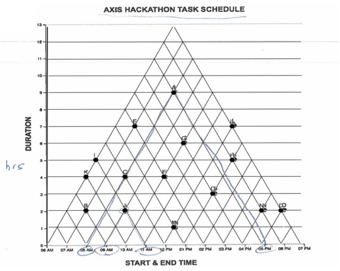


**Fig. 3a.** Pictorial intersections (Color figure online)

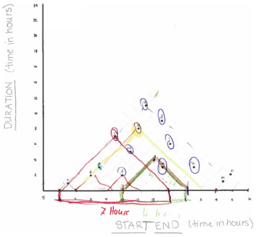


**Fig. 3b.** Annotations & examples (Color figure online)

In Fig. 3b (at left) participants have annotated their highlighted intersections. We see a partial worked example, via the annotation "7 h" to the span for the red interval.

In Fig. 3c (at right) we see a worked example where participants both highlighted the intersection and gave explicit values for a sample point on the plot. Under the graph they added a production rule for finding the start-time of a hypothetical point "S", indicating that some learners may prefer text instructions. *(triangular grid faded in digital scanning)*

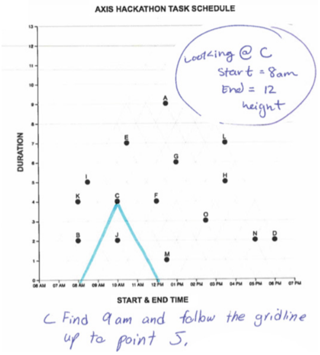


**Fig. 3c.** Worked example (Color figure online)

Finally in Fig. 3d (at right) we see explanatory text with an explicit definition of several graph elements.

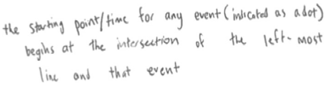

the starting point/time for any event (indicated as a dot)
begins at the intersection of the left-most
line and that event

**Fig. 3d.** Text instruction

### 2.3 Discussion of Study One

The results of Study One suggest the Triangular Model (TM) graph is challenging for STEM undergraduates. While the graph is elegant in its simplicity—as one participant noted, "once you see [the triangles], you can't (*sic*) unsee them"—most re-imagined the marks on the page as components of the more conventional representation for intervals. In interpreting this graph students invoked prior knowledge of conventions for the domain (intervals as line segments) and graphs in general (Cartesian coordinates). When prompted for instructional aids, students believed they could easily improve performance of future participants by adding instructions highlighting the multiple intersections of a point with the *x*-axis. Importantly, these scaffolds are substantively different than those explored in previous literature [2–6]. These instructions are most similar to graphical cues [3, 4], but rather than reinforcing the main argument of the graph (e.g. local maxima/minima, salient trend, etc.) they draw attention to the structure of the coordinate system. Both text and image instructions focus on the graphical framework and how to perform a first-order reading, rather than reinforcing the connection between the graph's signifiers and referents.

Owing to the limited sample size and observational methods, we fall short of explaining why some students (3 sessions) *were* able to interpret the graph while most were not. In one case, an individual interpreted the graph in the very first question, but failed to think-aloud, leaving their strategy a mystery. In the second case, the dyad pair also developed a correct model in the first question. In the third case, the dyad read the graph incorrectly for about half the questions before realizing their mistake and re-solving the problem set. These outcomes could be driven by individual differences in graphical competency, or different problem-solving strategies. Addressing this question will require further observation with directed post-task interviews.

## 3 Study Two: Testing Scaffolds for an Unconventional Graph

Inspired by the instructional aids produced by participants in Study One, we designed four scaffolds for self-directed learning: two text instructions (adjacent to the graphs) and two illustrations (highlighting *x*/*y* intersections). The "what-text" design (Fig. 4a) specifies the components of the graph and describes their meaning. The "how-text" design (4b) provides a set of production rules for extracting data from the graph. In the "static-image" (Fig. 4c), intersections are displayed for a single data point persistent throughout the task. In the "interactive-image" (Fig. 4d), the appropriate intersections appear & disappear when a participant hovers their mouse over any data point.

Prior work [11] has demonstrated that the computational efficiency of the TM graph can be achieved by students after 20 min of interactive video instruction. In Study Two we test the effectiveness of our designs by seeking to replicate these results with scaffolding alone. Assigning each participant to a scaffold condition, we compare their performance on both the LM and TM graphs, and subsequent ability to draw a TM graph for a small data set. We hypothesize that: (1) scaffolding will not affect performance on the LM graph, because it is conventional and relatively easy to read; (2) learners without scaffolding (control) will perform better with the LM than TM;

A point is an interval of time
The left intersection with the x-axis along the diagonal gridline is the start time
The right intersection with the x-axis along the diagonal gridline is the end time
The intersection with the y-axis is the duration.

**Fig. 4a.** "what-text" specifies graph components and their meaning

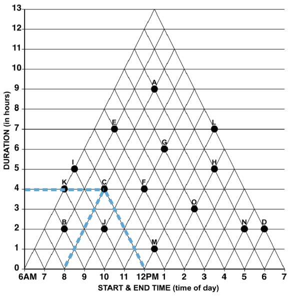


**Fig. 4c.** "static-image" displays *x/y* intersections for one data point

Start-time: follow the left-most diagonal gridline to the intersection with the x-axis
End-time: follow the right-most diagonal gridline to the intersection with the x-axis
Duration: follow the horizontal gridline to the intersection with the y-axis
Label: the letter directly above the point

**Fig. 4b.** "how-text" specifies how to extract data from the graph

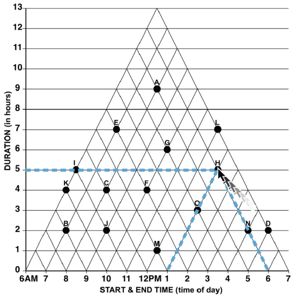


**Fig. 4d.** "interactive-image" displays *x/y* intersections on mouseover

(3) learners with (any form of) scaffolding will perform better with the TM than LM (replication of [11]). Finally, based on observations in Study One we expect that graph-order will act as a scaffold. (4) Learners who solve problems with the LM graph *first* will perform better on the TM (relative to TM-first learners) as their attention will be drawn to the salient differences between the graph types.

## 3.1 Methods

**Design.** We employed a 5 (scaffold: none-control, what-text, how-text, static image, interactive image) × 2 (graph: LM, TM) mixed design, with scaffold as a between-subjects variable and graph as a within-subject variable. To test our hypothesis that exposure to the conventional LM acts as a scaffold for the TM, we counterbalanced the order of graph-reading tasks (order: LM-first, TM-first). For each task, we measured response accuracy and time. For the follow-up graph-drawing task, we coded the type of graph produced by each participant.

**Participants.** 316 (69% female) STEM undergraduates aged 17 to 33 were recruited from the experimental-subject pool at a large American university (M(age) = 21, SD (age) = 2), yielding approximately 30 participants per cell in the 5 x 2 design.

**Materials**

*Scaffolds.* For the first five questions of each graph-reading task, participants saw their assigned scaffold along with the designated graph. On the following ten questions, the scaffold was not present. Examples of each scaffold-condition for the TM graph are shown in Fig. 4. Equivalent scaffolds were displayed for the LM graph (see footnote 1).

*The Graph Reading Task.* Each graph reading task consisted of a graph (LM or TM) and 15 multiple choice questions (used in Study One). Questions were presented one at a time, and participants did not receive feedback as to the accuracy of their response before proceeding. The order of the first five (scaffolded) questions was the same for each participant, while the order of the remaining 10 were randomized. For each question, the participant's response accuracy (correct, incorrect) and latency (time from page-load to "submit" button press) was recorded. Because each participant completed the reading task once with each graph, we developed two matched scenarios: a project manager scheduling tasks (scenario A), and an events manager scheduling reservations (scenario B). In each scenario, an equivalent question can be identified in the other pertaining to the same interval property/relation. For example, in scenario A the question mapping to the "starts" property reads: "Which tasks are scheduled to start at 1 pm?", and the correct answer consists of 2 tasks (Fig. 5 – *left* – tasks O & H). In scenario B, the equivalent question reads: "Which reservations start at 8:00 AM?", the correct answer referencing 3 events (Fig. 5 – *right* – events D, C & L). For the LM graphs, intervals were sorted in order of duration, with the longest appearing at the top of the graph. A pilot study on Amazon Mechanical Turk using the LM graph revealed no significant differences in response accuracy or latency between the scenarios. The four graphs constructed for the study are shown in Fig. 6.

*The Graph Drawing Task.* Participants were given a sheet of isometric dot paper with a table of 10 time intervals, and directed to draw a triangular graph of the data ("like the triangle graph you saw in the previous task"), using the pencil, eraser and ruler provided. Isometric dot paper equally supports the construction of lines at 0, 45 and 90 degrees, minimizing any biases introduced by the paper on the features of the graph drawn by participants.

**Procedure.** Participants completed the study individually in a computer lab. They completed the two graph-reading tasks in sequence, one with a TM graph and the other with an LM graph (order counterbalanced). Afterward, participants completed the graph drawing task. The entire procedure ranged from 22 to 66 min.

### 3.2 Results: The Graph Reading Task

Performance on graph-reading tasks is a combination of response accuracy (score) and time. Table 1 displays the mean values for score (as % correct) and time (in minutes) for each graph across the scaffold conditions. As we found little variance in response

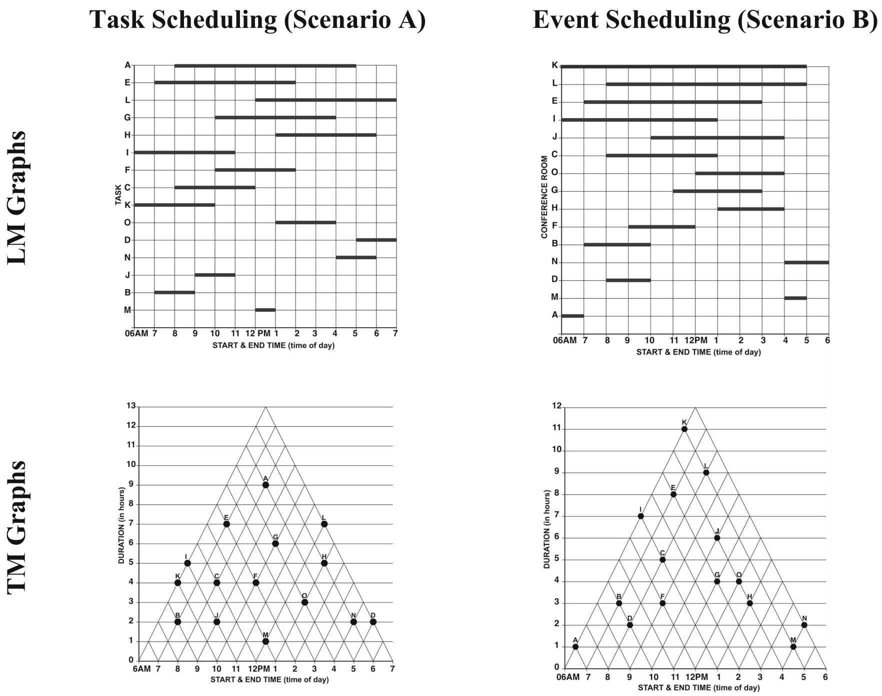


**Fig. 5.** LM and TM graphs for each scenario of graph reading task

**Table 1.** Mean score and response time for graph reading tasks

| | Mean score (% correct) | | | | Mean time (mins) | | | |
|---|---|---|---|---|---|---|---|---|
| | *LM graph* | | *TM graph* | | *LM graph* | | *TM graph* | |
| Scaffold | M | SD | M | SD | M | SD | M | SD |
| none-control | 73 | 16 | 46 | 30 | 8.6 | 2.1 | 11.2 | 3.6 |
| what-text | 74 | 15 | 59 | 29 | 9.8 | 2.9 | 11.6 | 3.6 |
| how-text | 73 | 14 | 58 | 31 | 9.1 | 2.3 | 10.9 | 3.0 |
| static-image | 73 | 15 | 57 | 30 | 9.1 | 2.6 | 10.6 | 3.5 |
| interactive-img | 73 | 13 | 71 | 23 | 9.4 | 2.6 | 9.9 | 2.6 |
| Total | 73 | 14 | 59 | 30 | 9.2 | 2.5 | 10.9 | 3.3 |

time we focus our discussion on performance as judged by response accuracy. To explore the potential influence of graph, scaffold, and graph-order on scores, we performed a mixed effects ANOVA on score with graph as the within-subjects factor, and scaffold, graph-order and scenario as between-subjects factors (Fig. 6).

**Effect of Graph.** We found a significant main effect of graph type on score, F (1,297) = 97.67, p < .001. In Fig. 6 we see that across all factors, LM scores [green

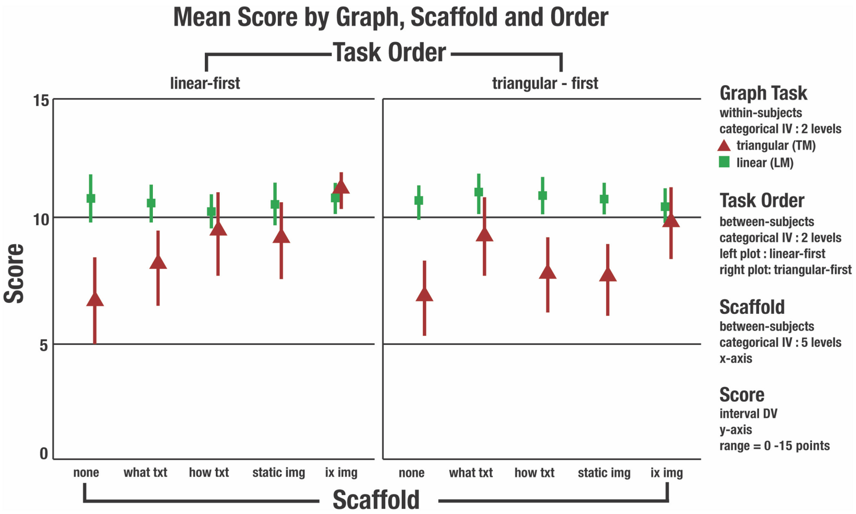


**Fig. 6.** Mean response score by graph, Scaffold and task order LM scores (squares) remain steady across scaffold (x-axis) and graph-order (right/left plot), while TM scores (triangles) differ by scaffold, highest in the interactive image condition. (Color figure online)

squares] (M = 10.96, SD = 2.13) are significantly higher than TM scores [red triangles] (M = 8.78 SD = 4.44), t(316) = −9.45, p < 0.001, r = 0.47, consistent with our motivating assumption that the TM graph is more challenging to interpret.

**Effect of Scaffold.** We found a significant main effect of scaffold on score, F(4,297) = 4.24, p < .05. A post-hoc t-test supports our second hypothesis, that across all other factors, participants in the no-scaffold control group performed significantly better with the LM graph (M = 10.98, SD = 2.33) than the TM graph (M = 6.9, SD = 4.51), t(60) = 7.07, p < 0.001, r = 0.67. Regarding our third hypothesis, we found a significant interaction between graph and scaffold, F(4,297) = 10.03, p < .001. As predicted, scaffolds did *not* influence the score when solving problems with the LM (hypothesis 1), but made significant improvements in score for the TM. However, *none* of our scaffolds resulted in significantly higher scores for the TM relative to the LM. In fact, post-hoc pairwise comparisons (with Bonferroni correction) on TM scores showed

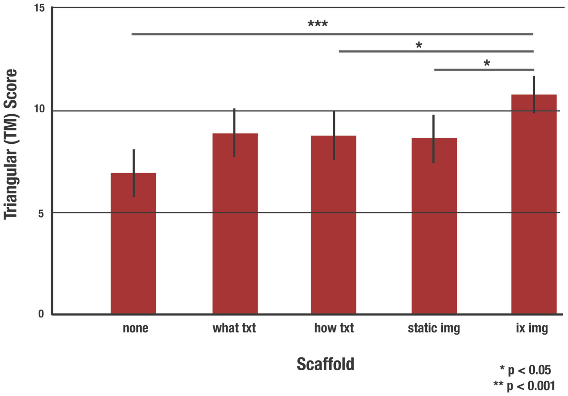
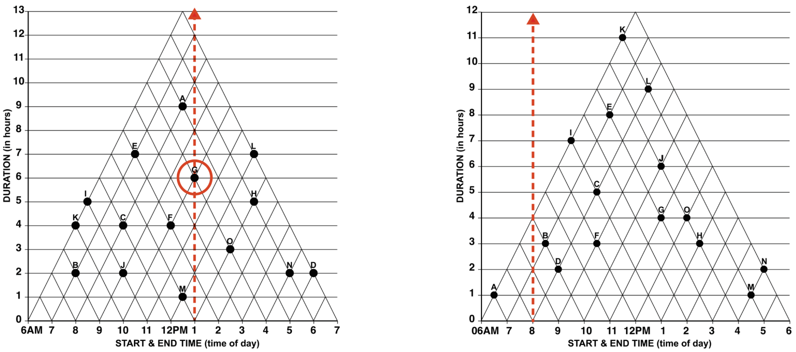


**Fig. 7.** Only the interactive image scaffold was significantly better than no-scaffold control condition.

that *only* the interactive image scaffold yielded scores significantly higher than the no-scaffold control (Fig. 7).

**Effect of Graph-Order.** Counter to hypothesis 4 that graph-order would act as a scaffold for comprehension, we found no main or interaction effects for graph-order on response accuracy. Perhaps in order to glean salient differences between the TM and LM graphs, they need to presented simultaneously (as in Fig. 1).

**Effect of Scenario.** As our mixed design necessitated the use of two matched scenarios, we tested for effects of scenario in our statistical model. Unexpectedly, we found a main effect of scenario on score, F(1,297) = 22.29, p < .001, and significant interaction between graph and scenario, F(1,297) = 34.34, p < .001. When answering questions in the "task scheduling" scenario A (M = 9.20, SD = 4.12), participants had significantly lower scores, t(316) = –4.77, p < 0.001, r = –.26, compared to the "events scheduling" scenario B (M = 10.52, SD = 2.97). In an online pilot we found no significant differences in performance between the scenarios when tested with the LM graph. To explore the source of this effect, we examined the data sets constructed for each scenario, and in particular, the very first question students solved with the TM graph. In the "task scheduling" scenario A (Fig. 8–*left*) we see that if a learner makes the most common mistake—seeking an orthogonal intersection from the *x*-axis—there is a single data point that intersects the line: an available answer. However, in the "events scheduling" scenario B (Fig. 8—*right*), there is **no** intersecting data point. Students who were randomly assigned to this second scenario received implicit feedback that they were misreading the graph if they sought the orthogonal intersect because there was no answer to the question. We suspect this drove students to re-evaluate their strategy, yielding significantly higher scores for the "events scheduling" scenario.



What tasks start at 1pm? A data point intersects the erroneous orthogonal projection.

What events start at 8 am? No data points intersect the mistaken erroneous projection.

**Fig. 8.** First question for the task (left) and event (right) scenarios.

### 3.3 Results: The Graph Drawing Task

The graph drawing tasks allows us how to explore how each scaffold supports students learning the graphical framework of the TM. We expect that accurately drawing requires deeper understanding of how the graph works, and analysis of any systematic mistakes students make in drawing may reveal sources of difficulty in comprehension. Following the directed approach to qualitative content analysis [12], a team of 3 raters classified all 316 drawings first into *a priori* categories [triangular, linear, other] and finally into 5 categories based on the data present in the sample: (correct) triangular, linear, scatterplot, "asymmetric triangular" and "right-angled". Interrater reliability was high ($\alpha = 0.96$) and disagreements were resolved through negotiation. The majority (73%) of participants drew correct TM graphs. 17 individuals (5%) constructed LM graphs, while 3 participants drew scatterplots with start & end time on the *x*/*y* axes respectively. Most interesting were the two alternative triangular forms constructed by 66 (21%) individuals: right-angle triangle, and asymmetric triangles (described in Fig. 9).

While the overall distribution of graph drawing-types was too heterogeneous to reliable correlate with TM task performance, we did examine the performance of the subset of participants who produced the two alternative triangular forms. TM scores for participants who drew "right-angle" graphs were significantly lower (M = 2.3, SD = 1.98) than for participants who drew "asymmetric triangle" graphs (M = 8.55, SD = 3.73), t(27.11) = –7.36, p < 0.001, r = 0.82.

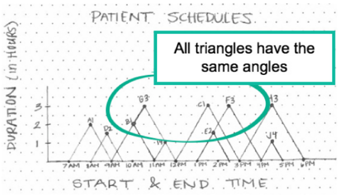


**Fig. 9a.** 230 students drew correct TM graphs

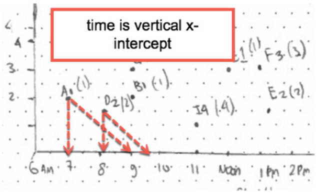


**Fig. 9c.** 44 students drew "right-angle" graphs. They plot duration on the Y axis and the interval as a point, but mistakenly use an orthogonal x-intersect for start time

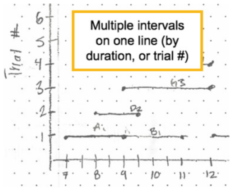


**Fig. 9b.** 17 students drew LM graphs

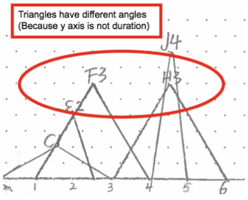


**Fig. 9d.** 22 students plotted the vertical intersection as the midpoint of the interval, but the triangles were not geometrically similar because duration was not on the y-axis.

### 3.4 Discussion of Study Two

The results of Study Two leave us with a conundrum: why were the scaffolds designed by learners in Study One largely ineffective?

None of our designs replicated the results of Qiang et al. [11] which yielded better performance with the TM than LM graph, though there were notable differences in our tasks, including their use of an interactive graph interface with hundreds of data points, and feedback in the video instruction. Setting aside the differences in performance *between* the LM and TM graphs, we assessed the efficacy of scaffold designs by looking at TM scores alone. The widely-held assertion of Study One participants that simple text and image instructions would dramatically improve readability of the graph were not borne out, as on average, participants who received the static scaffolds performed no better than those who received (as participants in Study One) no graph instructions at all (Fig. 7).

We suspect the source of this discrepancy lies in a hindsight bias. Once students understand how the graph works, they cannot "unsee" it, and therefore underestimate the strength of their prior expectations. The unexpected effect of scenario on TM scores supports this interpretation, as students who received implicit feedback they were reading the graph incorrectly (because there was no available answer) performed better than those who did not (Fig. 8 right vs. left). In this way, the structure of the task presented the reader with a *mental impasse* [13] where their expectations (based on prior knowledge of Cartesian graph forms) left them with no solution, and their attention was actively redirected to reconsidering these expectations. The role of attention can also address why the interactive image was superior to the static text and image scaffolds. If it is the case that a reader does not *realize* they are misreading the graph (as we observed in Study One), it is easy to ignore the static scaffolds. However, it is much more difficult to ignore a stimulus that appears every time the mouse is moved over a data point. To critically evaluate the role of attention in our ongoing studies we are employing both mouse and gaze-tracking to quantify the extent and time-course of attention paid to both scaffolds and graph features.

As in Study One, the most substantial open question in this work remains the source of individual differences. Across all conditions, we see a high standard deviation (30% or 5 points) in score, again with some participants in the no-scaffold control able to correctly interpret the graph. In our ongoing work we seek to address this question with post-task interviews that prompt participants to explain their interpretation strategy while viewing a screencast replay of the their problem-solving session.

## 4 General Discussion

While the Triangular Model (TM) graph is elegant in its simplicity, the results of our studies demonstrate this simplicity is *deceptive.* Without assistance, most readers misinterpret the graph as the conventional representation for time intervals: the linear model. Even with cognitive aids, many students persisted in this erroneous interpretation, and only an interactive image scaffold significantly improved comprehension.

These results have implications for both the design of scaffolds and of unconventional graphs. First, when designing scaffolds one should consider the reader's expectations based on the conventional representation for variables in the domain. It is from that prior knowledge that readers begin their interpretation, *not* from a blank-slate (i.e. general graph schema) we might expect based on a graph's surface features. To overcome this, our results suggest that techniques actively directing attention to salient differences may prove most effective. The interactive-image scaffold achieves this through repeated, user-driven exposure to the multiple intersections of a TM data point with the $x$-axis. Similarly, the *mental impasse* provided by the questions in our event-scheduling scenario actively directed readers' attention to their mistaken interpretation. We are presently conducting a follow-up study testing the relative efficacy of attention-directing explicit (e.g. interactive image) and implicit (e.g. mental impasse) scaffolds.

When constructing unconventional graphs, a designer's priority is the computational affordances making the new graph-form suitable to the data and task. But as we learn from these studies, a designer should also ask, "What expectations will be invoked by the marks on the page?" For the TM graph, we suspect it is the orthogonal axes that drive readers to expect a single orthogonal intersection for each data point. But there is—strictly speaking—no reason that the axes *need* to be orthogonal. In fact, one clever participant in our graph drawing task produced what we believe to be a substantial improvement upon the TM graph, where the $y$ axis was positioned diagonally on the left side of the graph's "bounding triangle". We are presently conducting a follow-up study to investigate alternative axis and grid designs, hypothesizing that such diagonally positioned axes will yield significantly better performance.

In this work, we have explored only a small subsection of the total design space of scaffolding techniques for a particular kind of unconventional graph. We expect our conclusions generalize to unconventional coordinate systems, but that other techniques need to be explored when employing unconventional *markings*. Our choice of scaffolds was inspired by direct observation and participatory design, however, we suspect a wider range of techniques might be effective in more instructional settings, including explication of worked examples, or seeing the graph being *drawn.* While we chose to separate our text and image scaffolds to test their differential efficacy, a combined text/image annotation could prove effective even in static media, and is a part of our ongoing work.

We started by reasoning that existing scaffolding techniques would be insufficient for unconventional graphs because learners would lack the prior knowledge of the new graph system required to make use of them. As Pinker [7] suggests, when confronted with an unfamiliar graph form, the reader instantiates a generic "general graph schema". However, it seems that despite differences in surface structure, a learner's prior knowledge of *other* graph forms can actively interfere with interpretation of a new graph. The novelty of the diagonal gridlines in the TM graph was not enough for most learners to suspend their Cartesian expectations. To overcome this prior knowledge, we think that successful scaffolds for unconventional graphs must not only show or tell us how to read them, but to rather alert us that that we need to pay attention, and reconsider our expectations in the first place.

**Acknowledgements.** Sincerest thanks are offered to research assistants Evan Barosay, Kai-Yu Chang and Hazel Leung, as well as Drs. Seana Coulson, David Kirsh, Rafael Núñez and Caren Walker.

## References

1. Larkin, J., Simon, H.: Why a diagram is (sometimes) worth ten thousand words. Cognit. Sci. **99**, 65–99 (1987)
2. Shah, P., Hoeffner, J.: Review of graph comprehension research: implications for instruction. Educ. Psychol. Rev. **14**(1), 47–69 (2002)
3. Acartürk, C.: Points, lines and arrows in statistical graphs. In: Cox, P., Plimmer, B., Rodgers, P. (eds.) Diagrams 2012. LNCS (LNAI), vol. 7352, pp. 95–101. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31223-6_13
4. Acartürk, C.: Towards a systematic understanding of graphical cues in communication through statistical graphs. J. Vis. Lang. Comput. **25**(2), 76–88 (2014)
5. Kong, N., Agrawala, M.: Graphical overlays: using layered elements to aid chart reading. IEEE Trans. Vis. Comput. Graph. **18**(12), 2631–2638 (2012)
6. Mautone, P.D., Mayer, R.E.: Cognitive aids for guiding graph comprehension. J. Educ. Psychol. **99**(3), 640–652 (2007)
7. Pinker, S.: Theory of graph comprehension. In: Freedle, R. (ed.) Artificial Intelligence and the Future of Testing, pp. 73–126. Erlbaum, Hillsdale (1990)
8. Shah, P., Freedman, E.G., Vekiri, I.: The comprehension of quantitative information in graphical displays. In: Shah, P., Miyake, A. (eds.) Cambridge Handbook of Visuospatial Thinking (2005)
9. Kulpa, Z.: A diagrammatic approach to investigate interval relations. J. Vis. Lang. Comput. **17**(5), 466–502 (2006)
10. Qiang, Y., Delafontaine, M., Versichele, M., De Maeyer, P., Van de Weghe, N.: Interactive analysis of time intervals in a two-dimensional space. Inf. Vis. **11**(4), 255–272 (2012)
11. Qiang, Y., Valcke, M., De Maeyer, P., Van de Weghe, N.: Representing time intervals in a two-dimensional space: an empirical study. J. Vis. Lang. Comput. **25**(4), 466–480 (2014)
12. Hsieh, H.-F., Shannon, S.E.: Three approaches to qualitative content analysis. Qual. Health Res. **15**(9), 1277–1288 (2005)
13. Ohlsson, S.: Information-processing explanations of insight and related phenomena. In: Advances in the Psychology of Thinking, pp. 1–44 (1992)

# Rediscovering Isotype from a Cognitive Perspective

Günther Schreder(✉), Nicole Hynek, Florian Windhager,
and Eva Mayr

Danube University Krems, Dr.-Karl-Dorrek-Str. 30, 3500 Krems, Austria
{guenther.schreder, nicole.hynek, florian.windhager,
eva.mayr}@donau-uni.ac.at

**Abstract.** Almost 100 years ago, Otto Neurath developed the Isotype (International System of Typographic Picture Education) method to communicate statistical information to the broad public in an intuitive, pictorial way. It translates numerical data into arrangements of repeated pictograms. This method is still well-used in information design and data journalism. Neurath's original publications contained a lot of assumptions on how Isotype diagrams are processed by recipients: e.g. they can be understood easily, because pictograms are processed in the same way as everyday observations of the same concepts. But documented empirical proof was entirely missing. We present a model for the reception of Isotype-like diagrams from a cognitive perspective. This model includes Isotype's positive effects of countability, iconicity and ancillary semantic information on graph comprehension. Positive effects on engagement and perceived attractiveness are included as additional factors commonly attributed to Isotype. We discuss existing empirical studies, point out research gaps and propose a roadmap for further research.

**Keywords:** Isotype · Information visualization · Pictographs
Pictorial representation

## 1 Introduction: Isotype as Conceptualized by Otto Neurath

In the early 1920s, Otto Neurath (1882–1945) with his team of designers, statisticians and transformers created the International System of Typographic Picture Education (Isotype). The main idea behind Isotype was to communicate statistical data on social, economic, and political topics to the broad public in an intuitive, pictorial way.

### 1.1 Design of Isotype

The method established a set of rules for a consistent design of pictorial statistics (Hartmann 2006a, b): Icons should (1) be used consistently for the same concept, (2) be of the same size, and (3) bear a strong resemblance to the object they represent ("speaking symbols", O. Neurath 1926/2006, p. 8). Contrary to many other visualizations at that time that used stretched icons (Fig. 1, top left), the icons of an Isotype visualization are countable and each of them stands for a concrete number of the respective concept

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 458–469, 2018.
https://doi.org/10.1007/978-3-319-91376-6_41

(as defined in the legend). The icons are repeated from left to right according to reading direction (Neurath 1936). Icons can be compared across years or countries on a vertical axis (Fig. 1, bottom left). Additionally, correlations can be shown by combining two or more different symbols (Fig. 1, right). Details should only be given if they have some informative value and Neurath argued that the actual numbers should not be presented in detail, as he was convinced that it is better for people to forget the actual numbers and instead remember the whole picture (Neurath 1936).

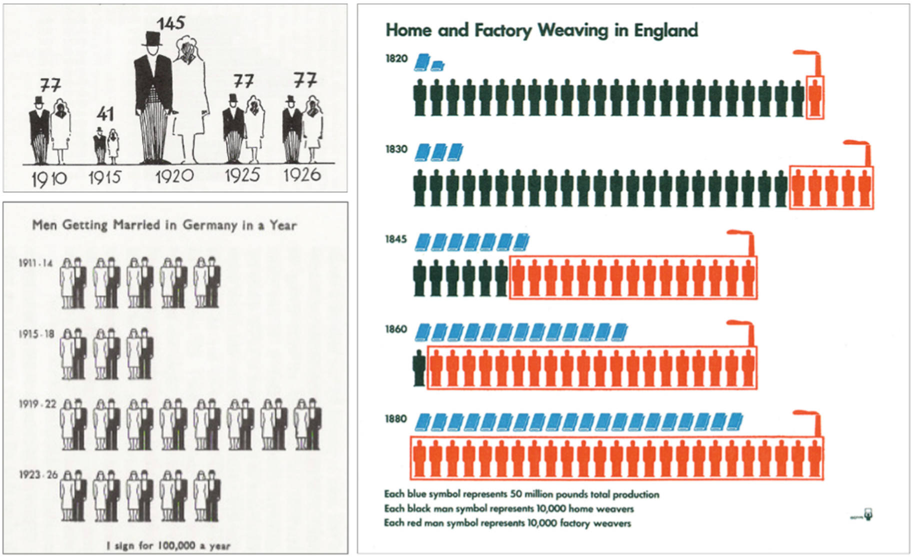


**Fig. 1.** NOT Isotype (top left) and Isotype visualization (bottom left) as explained by Neurath (1936), correlating variables in Isotype (right, Neurath 1939)

Even though the basic elements of the Isotype grammar are simple, the transformation process is not. In his autobiography, Neurath (2010) compares it to writing a novel: Knowing only the words and grammar will not make you a good author, “one also has to know how to select combinations of words to produce a striking result” (p. 102). This is highly evident when MacDonald-Ross (1977) compares an Isotype-like visualization by Vernon (1946) with an Isotype visualization of the same data by Marie Neurath (see Fig. 2): Using icons is not enough, their arrangement, the selection of data, and the guiding pictures are also important factors. In Neurath’s version critical information is added in the caption, whereas the labels are simplified. The pictograph-to-number-ratio is reduced and some years were omitted. As a result, Neurath’s visualization implies a specific message, i.e. a strict difference in the importance of military and civil occupation.

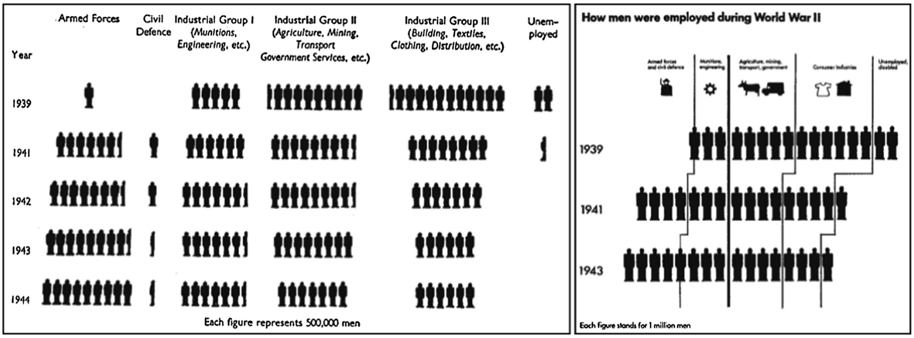


**Fig. 2.** A visualization of the same data by Vernon (1946, left) and Marie Neurath (right). Images were taken from MacDonald-Ross (1977).

### 1.2 Effects of Isotype

Otto Neurath wanted to communicate statistical facts in a way that made them readily comprehensible and easy to remember. The content of an Isotype visualization, the general view, should be recognizable at first sight (Neurath 1936). To facilitate this effect, Isotype builds on pictorial icons: According to Neurath, pictures are not only more attractive than numbers only, but these “speaking symbols” also share some of the visual attributes of the objects they represent and are thus easier to understand than the more abstract spoken or written language (Neurath 1936, p. 20). Furthermore, “Isotype symbols have fewer positive or negative associations than the words of a language” (Neurath 2010, p. 125). On controversial problems “both sides could get their arguments from the same chart” (Neurath 2010, p. 125). Neurath and his collaborators believed that Isotype presented neutral data and kept evaluation and judgment to the viewer (Neurath and Kinross 2009, p. 26).

Otto Neurath reported different evaluations of Isotype: Allegedly, psychological studies compared Isotype material with stretched icons with the result that “the Isotype technique is better” (Neurath 2010, p. 115). Additionally, his team conducted field observations in elementary schools and at the Museum of Society and Economy in Vienna, which helped to gain a more comprehensive understanding of the Isotype technique (Neurath and Kinross 2009). However, no detailed results are reported from any of these empirical approaches.

Although remnants of Neurath’s method are still present in contemporary infographics and data visualizations, his claims on their benefits have rarely been scrutinized. As there are no historical documents on the actual effectiveness of Isotype and newer research on the cognitive aspects of Isotype-like visualizations is rare, we want to review the Isotype principles from a cognitive-scientific perspective to close this gap.

## 2 A Cognitive Model of Isotype Reception

Within the existing publications on Isotype, different effects have been described, which can be summarized in cognitive terms as follows: (1) Pictograms in Isotype can be counted, making it easier to grasp the numerical information than in non-countable graphs (*countability*), (2) the iconic character of pictograms assists the comprehension process (*iconicity*), (3) additional information like title, caption or guiding pictures assist the comprehension process (*ancillary semantic information*), and (4) the increased attractiveness of Isotype pictures results in higher engagement *(aesthetic preference).* In the following, we review existing research from cognitive science and information visualization and discuss the most relevant results for each aspect.

## 2.1 Countability Effect

As a general rule, Neurath proposed to reduce large numbers to a small amount of discrete objects, each representing a fixed quantity to facilitate the comprehension of statistical information. Does recent research support this idea?

One possible starting point is the phenomenon called *subitizing*, our ability to immediately generate exact counts of a small number of objects with highest precision (Haroz et al. 2015). Subitizing has been observed with up to four objects and is hypothesized to stem from a specific mechanism of the visual system (Choo and Franconeri 2013). In their first two experiments, Haroz et al. (2015) found substantial evidence for a beneficial effect of subitizing on the perception of Isotype visualizations. Participants were shown four different variations of visualizations (Fig. 3) for 1.5 s and were asked to write down the shown numbers immediately afterwards. Stacked visualizations significantly reduced the observed error rates in comparison to stretched visualizations, but only in the range of 1–5 elements. This effect vanished for higher numbers. There was no difference between abstract shapes and pictographs. Haroz et al. (2015) assume that the redundant encoding of the data as length as well as a small countable numbers leads to an improved memory performance.

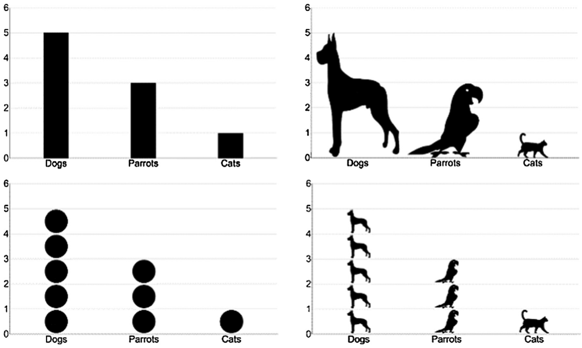
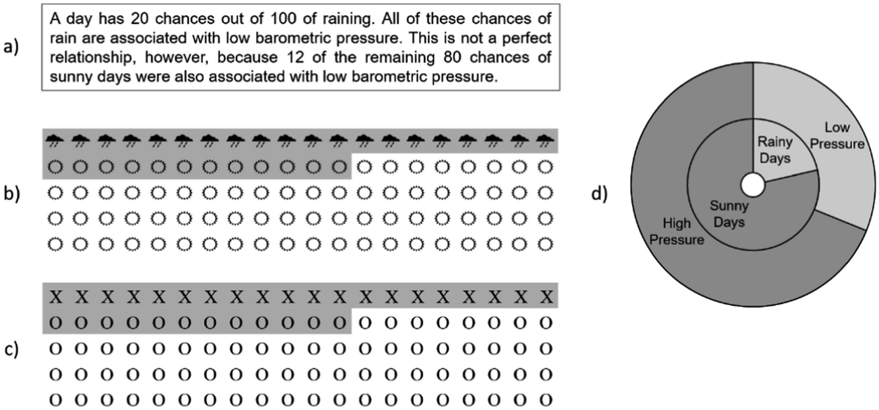


**Fig. 3.** The four variations used in Haroz et al. (2015): stretched vs. stacked and shapes vs. pictographs, (source: http://steveharoz.com/research/isotype/)

While it is widely accepted that graphical representations can promote understanding and improve statistical reasoning (McDowell and Jacobs 2017), theories of underlying causes are highly controversial and are commonly tested empirically in the field of Bayesian reasoning. Brase (2014) studied whether our brain is especially well-suited to deal with frequencies because it evolved to deal with the discrete objects of the real world (frequency hypothesis), or if any representations that provide transparent information on the structure of a subject matter would improve performance (nested sets hypothesis). In four experiments, he gave participants a Bayesian reasoning task with varying pictorial representations. Participants in the condition with icons as visual aids outperformed the participants in the other conditions (roulette wheel representation or no visual aid) by far (Fig. 4). Brase concludes that "these results indicate that icon representations, which better approximate actual ecological presentations of frequencies, are the most powerful pictorial technique currently known for facilitating correct Bayesian reasoning" (p. 93). But similarly to Haroz et al. (2015), no differences between realistic and abstract icons were found.



A day has 20 chances out of 100 of raining. All of these chances of rain are associated with low barometric pressure. This is not a perfect relationship, however, because 12 of the remaining 80 chances of sunny days were also associated with low barometric pressure.

**Fig. 4.** The conditions used by Brase (2014): the control group had text information (a) only, the experimental groups additionally had pictographs (b), abstract icons (c), or roulette wheel representation (d)

Though we do not have ample evidence on a countability effect of Isotype, recent research at least indicates that memory performance and reasoning can be improved through the visualization of discrete icons. Perhaps Neurath was even aware of one of the underlying mechanisms: According to Revkin et al. (2008), first accounts on subitizing date back as far as 1908 (Bourdon 1908).

## 2.2 Iconicity Effect

Expectations for actual effects of figurative pictographs on the perception process are very high: For example, Tversky (2011) suggests likenesses (i.e., pictorial representations of objects) to be readily recognized, more quickly understood, and better memorized. It has been proposed that likenesses could possibly support the decoding of graphs: "If the form of the external representation matches the internal form of the mental representation the workload for the cognitive system gets minimized"

(Rehkämper 2011, p. 4112). A low cognitive load can be desirable for perception, as it allows recipients to discern the topic and the main information quickly by reducing the mental effort necessary to decipher detailed data. A lowered cognitive load induced by easy perception also means that resources can be redirected to reflection and reasoning, i.e. from elementary to higher levels of graph comprehension (Friel et al. 2001). Nevertheless, the experiments mentioned above did not find any differences for fully abstract visual marks (like geometric shapes) and figurative icons. Is the iconicity-effect only a myth?

There is empirical data that points to positive effects of iconicity, but the conditions that lead to observable effects have to be scrutinized. In their third experiment, Haroz et al. (2015) found that visualizations featuring pictographs lead to a slightly lower error rate using a 1-back design. This means that participants were tested on the chart they saw before the currently presented chart and thus had to memorize both to be able to answer correctly. Haroz et al. (2015) conclude that pictographs should be used for demanding tasks as the information is recalled more accurately when working memory is under load.

Lin et al. (2013) studied whether bar-charts with semantically-resonant colors improve the speed of proportional comparisons by asking participants questions like "Which is larger, blueberry + peach or grape + banana?" (Fig. 5). Compared to a regular color assignment, response times were somewhat smaller for semantically-resonant charts.

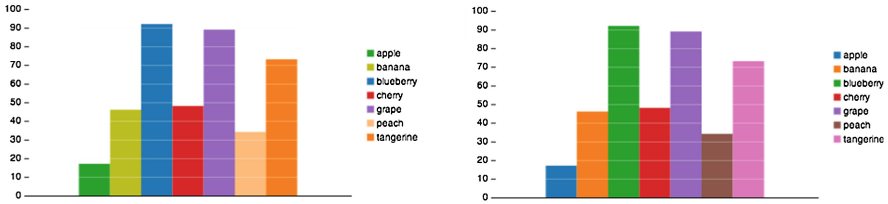


**Fig. 5.** Semantically-resonant colors (left) and standard palette (right) as used by Lin et al. (2013) (Color figure online)

Though this experiment did by no means aim at pictographs we argue that it taps into a similar cognitive process: Semantically-resonant colors not only help to discriminate between categories but also help to relate the numerical data to meaningful objects. As the majority of time spent in graph comprehension is spent on reading and reexamining axes and labels, Carpenter and Shah (1998) assume that people are often not able to keep this information in the working memory. Pictographs in Isotype as well as semantically-resonant colors (that were used in Isotype as well, cf. Neurath 1936) add this semantic information to the quantitative information in the graph, thereby possibly reducing the amount of information recipients have to hold in working memory.

The specific task as well as the seven categories used by Lin et al. (2013) and the 1-back design by Haroz et al. (2015) definitely put a bigger strain on the working memory than the studies on Bayesian reasoning (like Brase 2014), which mostly use only one or two different categories. In the latter case little can be won by using pictographs to support the cognitive process of relating the data to real world objects.

Pictographs can also aid memory on a more global level: Borkin et al. (2016) studied recognition and recall in 393 different visualizations. Participants viewed 100 randomly selected visualizations for 10 s each while their eye movements were tracked. Afterwards, the 100 selected visualizations and 100 new visualizations were presented for 2 s each in a random order. Participants had to indicate recognition of the visualizations that had been shown in the previous phase. Finally, blurred, smaller images of the target visualizations were presented and participants had to write down as many descriptions as possible. On average, the quality of descriptions was higher for the visualizations with pictograms. Pictograms received less visual attention (fixation time) than other parts of the visualization, but assisted recognition and recall of the target visualizations. One explanation by the authors is that redundant information (of the data or message) assists with understanding and recalling the visualization.

## 2.3 Ancillary Semantic Information

Original Isotype visualizations not only consisted of numerical data encoded in rows of pictographs, but also of a title and captions. The experiment by Borkin et al. (2016) provides strong arguments for this strategy: They found that during the encoding phase participants spent most time looking at text elements, especially titles, and that titles were the most prominent feature they described or even reworded during the recall phase.

Additionally, many Isotype visualizations had guidance pictures or icons that provided additional qualitative or contextual information (Fig. 6). From a contemporary perspective, these could easily be disregarded as mere embellishments without further purpose - what Tufte (1983) would have called “chart junk”. However, a minority of information visualization researchers highlight the informative value these elements can bear. In their design space on “figurative frames”, Byrne et al. (2017, p. 19) distinguish different roles of figurative elements in information visualization: They can provide a background context, show data content or label data points. Still, irrelevant images have no connection to the data and topic and will hinder comprehension and recall.

In their experiment on “useful junk” Bateman et al. (2010) compared plain bar charts to embellished charts created by Holmes (1984) for Time Magazine (Fig. 7). The visualizations were presented to the participants, who first had to describe the charts with as much detail as possible by answering the experimenter’s questions (e.g. “What is the basic trend of the graph?”). Half of the participants were asked to recall as many of the charts as possible immediately afterwards, the other half had to recall the charts two to three weeks later. Though no differences regarding the details of the description or the immediate recall were found, the participants were able to recall more details on the charts by Holmes in the long-term condition. Arguably, this result can be explained by elaborative encoding as the additional information provided by the embellishments provided a deeper level of processing.

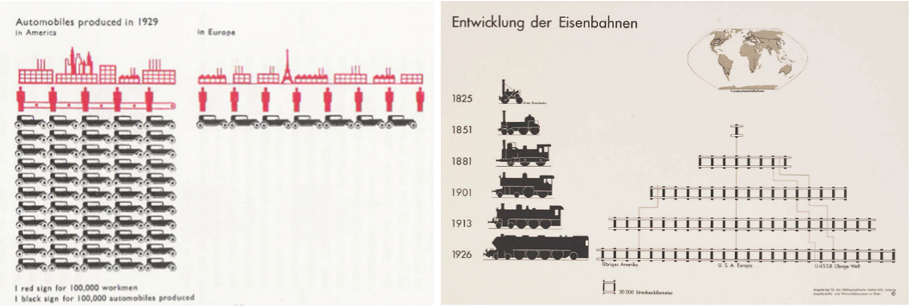


Fig. 6. Left: "Automobiles produced in 1929" (Neurath 1936) - guidance pictures on top provide additional cues on the categories. Right: "The Development of the Railroad" (Gesellschafts- und Wirtschaftsmuseum Wien 1930) - note that not only the length of the rails in different years is depicted and contextualized with a world map, but also the technological development of traction vehicles is shown as a series of pictographs.

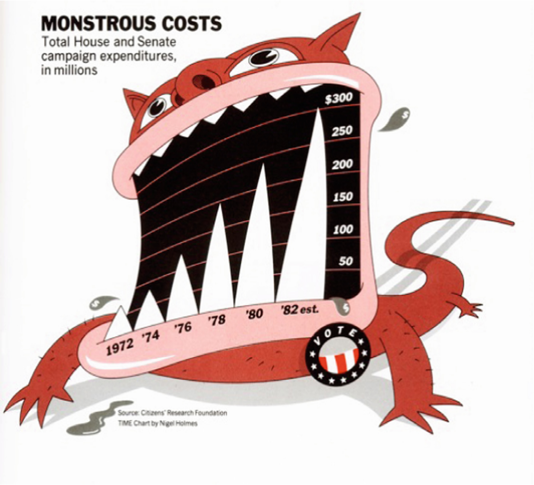


Fig. 7. Chart used by Bateman et al. (2010), provided by Nigel Holmes

## 2.4 Aesthetic Preference

Neurath (1936) argued that Isotype images are more attractive to recipients than numbers, but did not offer any explanations on potential causes. Overall, recent research supports his general presumption: For example, in their experiment on risk perception, Zikmund-Fisher et al. (2014) found higher preference ratings for charts with human icons or photographs in comparison to blocks and abstract representations of faces, especially for participants with lower numeracy and graphical literacy scores. Similarly, Gaissmaier et al. (2012) studied the reception of health-related statistical information presented in five different levels of iconicity (numerical information, abstract icons, pictographs, detailed pictographs, and photographs). Though they found no effect of the iconicity levels on memory, attractiveness was generally rated higher for graphical information than numerical information.

Haroz et al. (2015) studied how Isotype visualizations attract the participants' attention in comparison to other visualizations. Participants were able to choose between 9 different visualizations seen as a blurred preview to explore different topics. They were able to select a visualization, inspect it for as long as they wanted, and then return to the previews to select another visualization. Isotype visualizations did attract initial attention much more, and a higher engagement was consistently observed throughout the experiment.

To properly analyze the underlying processes of aesthetic preferences, Graf and Landwehr (2015) suggest to study more than generic liking judgments: According to dual process theories, automatic and controlled processes can be distinguished. Automatic processing is stimulus-driven and the result of a "passive exposure". It leads to a generic aesthetic judgement of pleasure or displeasure. In addition, controlled processes, which are dependent on the motivation to further process or interact with the stimuli, can lead to interest, boredom or confusion. Haroz et al. (2015) obviously aimed at the controlled processing, because the participants were able to actively choose the visualizations they wanted to inspect. As the motivation to engage with visual stimuli opens the variety of aesthetic evaluation beyond generic liking judgements, future research could focus on elements of Isotype that include stimulus-intrinsic triggers that increase the motivation to actively engage with the visualizations.

## 3 Discussion and Roadmap for Further Research

For a method that is nearly 100 years old and was widely used in social science and is still used in information visualization, research is extremely limited. Nevertheless, a few empirical studies – only one of them was directly aimed at analyzing Isotype – support some of Otto Neurath's claims. We tried to disentangle possible effects of Isotype into four categories (countability, iconicity, ancillary semantic information and attractiveness) and reviewed recent publications on these topics. Next, we want to discuss open questions and limitations of present research and propose future research questions.

**Which Effects of Isotype Are Supported by the Reviewed Research?** We found some evidence for benefits of using discrete icons that can be linked to the well-known phenomenon of subitizing. Still, comparisons with other well-established visualization methods are scarce and we do not know under which circumstances this method is actually more powerful than others. Subitizing suggests to only use a very low number of icons per variable, but could this effect possibly extend to differences in variables? Can differences be accurately determined by a single gaze when variables are depicted with a larger number of pictographs?

Iconicity effects are even more contested as most studies do not find any differences between abstract icons and pictographs. A possible reason is that pictographs only support the comprehension process when the working memory is under load, which is not the case in most experiments. Until now, this deduction remains purely a working hypothesis.

*Future experiments on iconicity should consider cognitive load in their experimental design. Presumably, tasks that are more strenuous for the working memory and dependent variables like time on task are required to test for iconicity effects.*

Otto Neurath stated that viewers should understand the most important facts from an Isotype picture at first sight. He wanted to communicate the main idea, but not concrete numbers. A rather pointed and often cited quote (e.g. Jansen 2009) that demonstrates Neurath's view is: "To remember simplified pictures is better than to forget accurate figures" (Neurath 1973, p. 220). Most of the reviewed studies used the accuracy of statistical problem solving as dependent variables (Brase 2009, 2014), but also recall of concrete numbers (Haroz et al. 2015; Zikmund-Fisher et al. 2014). These concepts have little in common with Neurath's idea. In contrast, gist knowledge as applied by Gaissmaier et al. (2012) refers to this idea of an overall understanding of the message. Also the research design applied by Borkin et al. (2016) is quite close to Neurath's idea: recognizing already seen visualizations and describing their contents in one's own words.

*For further research, we suggest to focus on more complex outcomes than problem solving: visual attention, gist knowledge and free recall are promising approaches. But further research should also go beyond comprehension processes: Otto Neurath highlighted a possible impact of Isotype on opinion formation. A recent study by Boy et al. (2017) investigated the influence of human icons on empathy. Though they did not find a positive effect, it would be worthwhile to also study outcomes on an emotional or motivational level.*

**Which Elements of the Isotype Design Space have been Studied?** Taking a closer look at the diagrams used in the experiments, we found that none of them follows the full Isotype grammar; rather we have to contend that the only commonality is that they use naturalistic icons. For example, Haroz et al. (2015) use vertical instead of horizontal alignment, and different studies on statistical reasoning (Brase 2009, 2014) arrange the icons in an icon array or Venn diagram. Neurath once stated that we should not turn "boring rows of numbers into boring rows of symbols" (Neurath and Kinross 2009, p. 104). Rather, he wanted to make more complex charts with more than one variable and present them in a way that triggers questioning and reflective thinking by the recipients.

*Therefore, it would be worth investigating not only the effect of using pictograms, but also some of the other design principles inherent to Isotype: for example, selection of data, more complex principles of arrangement, use of guiding pictures.*

Historically speaking, it is interesting to note that there is little research on a formerly important method. While most researchers agree that Isotype continues to influence contemporary InfoVis design, its fundamental principles remained largely untested and unverified. Overall, the reviewed studies show some promising results for Isotype visualizations, but more research into their effects and design principles is needed to give a final verdict on Otto Neurath's claims from a cognitive perspective.

The type of cognitive phenomena that can be investigated through experiments on information visualizations is largely dependent on the specific designs and visualization methods used to create test stimuli. Both the aesthetics and the aims of Isotype are very different than their counterparts in the majority of contemporary information visualization designs. Recently, information visualizations are more frequently regarded as everyday objects intended for laypeople, and not only tools for experts (Pousman et al. 2007) – which could be seen as a chance to revisit a method that was invented with similar goals? Based on our first insights, we suggest that a renewed focus on cognitive phenomena that have been neglected in the field of graph comprehension could profit from an investigation of this historical example.

## References

Bateman, S., Mandryk, R.L., Gutwin, C., Genest, A., McDine, D., Brooks, C.: Useful junk? The effects of visual embellishment on comprehension and memorability of charts. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2573–2582. ACM, New York (2010)

Borkin, M.A., Bylinskii, Z., Kim, N.W., Bainbridge, C.M., Yeh, C.S., Borkin, D., Pfister, H., Oliva, A.: Beyond memorability: visualization recognition and recall. IEEE Trans. Vis. Comput. Graph. **22**(1), 519–528 (2016)

Bourdon, B.: Sur le temps nécessaire pour nommer les nombres. Revue philosophique de la France et de l'étranger **65**, 426–431 (1908)

Boy, J., Pandey, A.V., Emerson, J., Satterthwaite, M., Nov, O., Bertini, E.: Showing people behind data: does anthropomorphizing visualizations elicit more empathy for human rights data? In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 5462–5474. ACM, May 2017

Brase, G.L.: Pictorial representations in statistical reasoning. Appl. Cogn. Psychol. **23**(3), 369–381 (2009)

Brase, G.L.: The power of representation and interpretation: doubling statistical reasoning performance with icons and frequentist interpretations of ambiguous numbers. J. Cogn. Psychol. **26**(1), 81–97 (2014)

Byrne, L., Angus, D., Wiles, J.: Figurative frames: a critical vocabulary for images in information visualization. Inf. Vis. (2017). https://doi.org/10.1177/1473871617724212

Carpenter, P.A., Shah, P.: A model of the perceptual and conceptual processes in graph comprehension. J. Exp. Psychol.: Appl. **4**(2), 75 (1998)

Choo, H., Franconeri, S.L.: Enumeration of small collections violates Weber's law. Psychon. Bull. Rev. **21**, 93–99 (2013). https://doi.org/10.3758/s13423-013-0474-4

Friel, S.N., Curcio, F.R., Bright, G.W.: Making sense of graphs: critical factors influencing comprehension and instructional implications. J. Res. Math. Educ. **32**, 124–158 (2001)

Gaissmaier, W., Wegwarth, O., Skopec, D., Müller, A.S., Broschinski, S., Politi, M.C.: Numbers can be worth a thousand pictures: individual differences in understanding graphical and numerical representations of health-related information. Health Psychol. **31**(3), 286 (2012)

Gesellschafts- und Wirtschaftsmuseum Wien: Gesellschaft und Wirtschaft: bildstatistisches Elementarwerk. Bibliographisches Institut, Leipzig (1930)

Graf, L.K., Landwehr, J.R.: A dual-process perspective on fluency-based aesthetics: the pleasure-interest model of aesthetic liking. Pers. Soc. Psychol. Rev. **19**(4), 395–410 (2015)

Haroz, S., Kosara, R., Franconeri, S.L.: Isotype visualization: working memory, performance, and engagement with pictographs. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 1191–1200. ACM, April 2015

Hartmann, F.: After Neurath: the quest for an inclusive form of the icon. In: Symposium Lecture at Stroom, Den Haag (2006a). http://www.medienphilosophie.net/artikel/

Hartmann, F.: Bildersprache [picture language]. In: Hartmann, F., Bauer, E.K. (Hg.) Bildersprache. Otto Neurath Visualisierungen, S. 15–105. Facultas Verlag, Wien (2006b)

Holmes, N.: Designer's Guide to Creating Charts and Diagrams. Watson-Guptill Publications, New York (1984)

Jansen, W.: Neurath, Arntz and ISOTYPE: the legacy in art, design and statistics. J. Des. Hist. **22**, 227–242 (2009)

Lin, S., Fortuna, J., Kulkarni, C., Stone, M., Heer, J.: Selecting semantically-resonant colors for data visualization. Comput. Graph. Forum **32**(34), 401–410 (2013). https://doi.org/10.1111/cgf.12127

Macdonald-Ross, M.: How numbers are shown. Educ. Technol. Res. Dev. **25**(4), 359–409 (1977)

McDowell, M., Jacobs, P.: Meta-analysis of the effect of natural frequencies on Bayesian reasoning. Psychol. Bull. (2017). https://doi.org/10.1037/bul0000126

Neurath, M., Kinross, R.: The Transformer. Principles of Making Isotype Charts. Hyphen Press, London (2009)

Neurath, O.: International Picture Language. The First Rules of Isotype. Kegan Paul & Company, London (1936)

Neurath, O.: Modern Man in the Making. Knopf, New York (1939)

Neurath, O.: From Hieroglyphics to Isotype: A Visual Autobiography. Hyphen Press, London (2010)

Neurath, O.: From Vienna Method to Isotype. In: Empiricism and Sociology, pp. 214–248. Springer, Dordrecht (1973)

Pousman, Z., Stasko, J., Mateas, M.: Casual information visualization: depictions of data in everyday life. IEEE Trans. Vis. Comput. Graph. **13**(6), 1145–1152 (2007)

Rehkämper, K.: Some remarks on pictorial statistics. In: Bulletin of the International Statistical Institute Proceedings of the 58th World Statistics Congress 2011, Dublin, pp. 4112–4114. International Statistical Institute, The Hague (2011)

Revkin, S.K., Piazza, M., Izard, V., Cohen, L., Dehaene, S.: Does subitizing reflect numerical estimation? Psychol. Sci. **19**(6), 607–614 (2008). https://doi.org/10.1111/j.1467-9280.2008.02130.x

Tufte, E.: The Visual Display of Quantitative Information. Graphics Press, Cheshire (1983)

Tversky, B.: Visualizing thought. Top. Cogn. Sci. **3**, 499–535 (2011)

Vernon, M.D.: Learning from graphical material. Br. J. Psychol. **36**, 145–158 (1946)

Zikmund-Fisher, B.J., Witteman, H.O., Dickson, M., Fuhrel-Forbis, A., Kahn, V.C., Exe, N.L., Valerio, M., Holtzman, L.G., Scherer, L.D., Fagerlin, A.: Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. Med. Decis. Mak. **34**(4), 443–453 (2014). https://doi.org/10.1177/0272989X13511706

# Perceptual Processing and the Comprehension of Relational Information in Dynamic Diagrams

Richard Lowe[1,2](✉), Jean-Michel Boucheix[1], and Marilyne Menant[1]

[1] University of Bourgogne Franche-Comté (LEAD-CNRS), Dijon, France
jean-michel.boucheix@u-bourgogne.fr
[2] Curtin University, Perth, Australia
r.k.lowe@curtin.edu.au

**Abstract.** To date, research on the processing involved in comprehending and learning from animated diagrams has accorded a minor role only to perceptual operations in general and peripheral processing in particular. For those aspects where the role of perception is acknowledged, it is foveal rather than peripheral processing that is regarded as the main player. In this paper, we use the results from additional finer grained analysis of data collected in a recent empirical study to suggest that information from a viewer's peripheral field can play a much more central role in animation processing than has previously been recognized. It appears that if the dynamic information comprising an animated diagram is presented in a suitable way, the resources available for visual perception can be partitioned so that responsibility is shared efficiently between foveal and peripheral processing. Implications with regard to elaboration of the Animation Processing Model and possible interventions for improving animation processing are discussed.

**Keywords:** Animated diagrams · Peripheral processing · Relation formation

## 1 Introduction

In this paper, we elaborate some aspects of the five-phase Animation Processing Model (APM) [1, 2], a theoretical framework concerned with the perceptual and cognitive activity involved in comprehending an externally presented animated diagram depicting complex, unfamiliar subject matter. The basic building blocks for this activity are individual event units (where event units are the entities depicted in an animation plus their associated dynamics). Figure 1 summarizes the APM phases involved in building a high quality mental model from an animation by a blend of bottom-up and top-down processes. APM Phase 1 processing involves a viewer's initial parsing to decompose the animated display's continuous flux of information into the separate event units that constitute the raw material for further processing. The APM characterizes this initial decomposition as an essentially bottom-up activity that is based largely on perceptual attributes of the animated display. The other four phases involve the progressive composition of event units into increasingly inclusive knowledge

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 470–483, 2018.
https://doi.org/10.1007/978-3-319-91376-6_42

structures that culminate in a high quality mental model of the depicted subject matter. For the purposes of this article, our discussion is confined to Phases 1, 2 and 3 because the role of perceptual processing is considered much less important in the cognitively-dominated 4th and 5th phases.

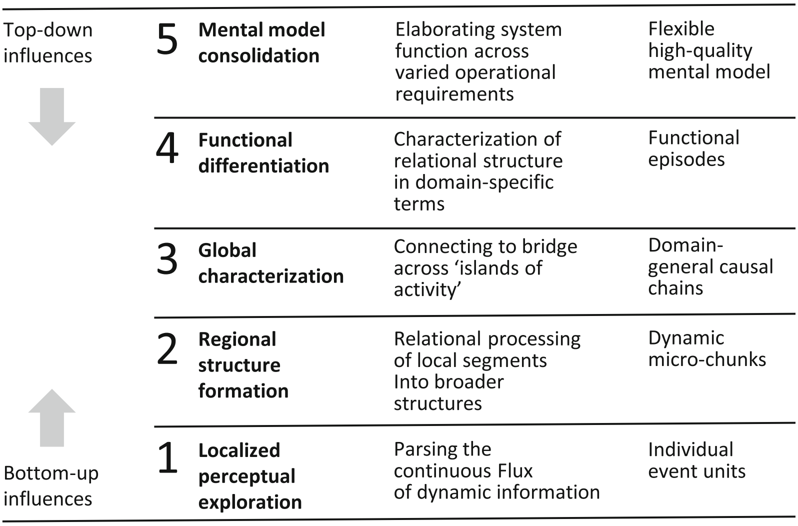


**Fig. 1.** The animation processing model summary diagram showing its five phases, the processing involved in each of the phases, and the outcomes of that processing.

With respect to bottom-up, perceptually-based activity, our initial version of the APM accorded a minor role only to peripheral processing and essentially confined its involvement to supporting Phase 1 decomposition. We implied that once event units had been separated out as raw material for mental model construction, the perceptual activity contributing to their subsequent hierarchical composition into causal chains essentially relied on foveal processing alone. This implication was consistent with the then prevailing orthodoxy regarding the centrality of such processing (a view probably reflecting the fact that measuring foveal fixations is the basis of eye tracking, a technique of proven value to the research community).

An example of the implied reliance on foveal processing can be illustrated by APM Phase 2 which involves the viewer's primary composition of adjacent event units into small local groups (termed *dynamic micro-chunks*). These groups are posited to be formed by the bonding together of multiple event units on the basis of domain general relationships, a knowledge of which has been acquired from the viewer's experience of the everyday world and its dynamics. Fundamental to the formation of dynamic micro-chunks are cause-effect relations where there is (i) a directed association between multiple dynamic changes that occur close together in time, and (ii) some type of link that connects these changes. Our initial exposition of the APM attributed viewer extraction of information about both the dynamic changes occurring (e.g., movements of components) and the interactions that linked adjacent components by cause-effect relations to processing involving solely foveal fixations.

The elaboration of the APM presented later in this paper was prompted by empirical and theoretical research undertaken in the ten years since its original exposition [1]. Although this work initially targeted APM Phase 1 processing (the preliminary decomposition of an animated display), the present contribution relates to subsequent activities in Phases 2 and 3 by which the viewer progressively composes individual event units into higher order knowledge structures. Our particular focus is upon the necessity of effective perceptual processing as a foundation for characterizing key relationships depicted in an animated diagram.

### 1.1 Composing Relations

The APM provides an account of the processing involved in learning from conventionally-designed animations that provide a comprehensive and dynamically faithful representation of their subject matter. On the basis of this account we hypothesized that deficiencies in learning from such materials are due to a mismatch between (i) the dominant design approach used to develop conventional animations and (ii) the way learners actually process dynamic representations [2]. If this is so, better alignment between animation design and learner processing should improve learning. A novel animation design was therefore devised that departed radically from the standard approach of providing a veridical presentation of the subject matter's dynamics. This alternative 'Composition Approach' [2] provides learners with a contiguous succession of carefully crafted partial (mini) animations designed to facilitate extraction of relevant information and its progressive composition into higher order mental structures. Using the working of a traditional upright piano mechanism as the to-be-learned subject matter (Fig. 2.), we employed three experimental conditions to compare the effectiveness of different animation designs (see [3] for experiment details): (i) Comprehensive (conventionally designed, with all components and behaviors included, as per Fig. 2), (ii) Contiguous (a succession of partial animations, each depicting two contacting and directly-relatable components at a time), and (iii) Non-contiguous (a succession of partial animations in which pairs of components were not in contact and therefore not directly relatable) (Fig. 3). The set of partial animations in both the Contiguous and Non-contiguous versions covered all components depicted in the Comprehensive animation.

In essence, the varied set of pivoted components comprising a piano mechanism act as two chains of interacting levers. When the pianist presses down on a piano key (lower right, Fig. 2), this input motion is transferred from lever to lever through the mechanism so that it finally reaches the output components that are directly responsible for producing the corresponding musical note (hammer, string) then stopping the note from sounding (damper). The intervening levers between input and output perform vital roles in controlling and coordinating these actions, particularly the proper sequencing of the hammer and damper movements. When a note is played, the various component motions required to produce the piano's proper functioning occur simultaneously or in rapid cascades. This makes its operation appear very complex to the uninitiated and so challenging to learn. Participants in the piano animation experiment

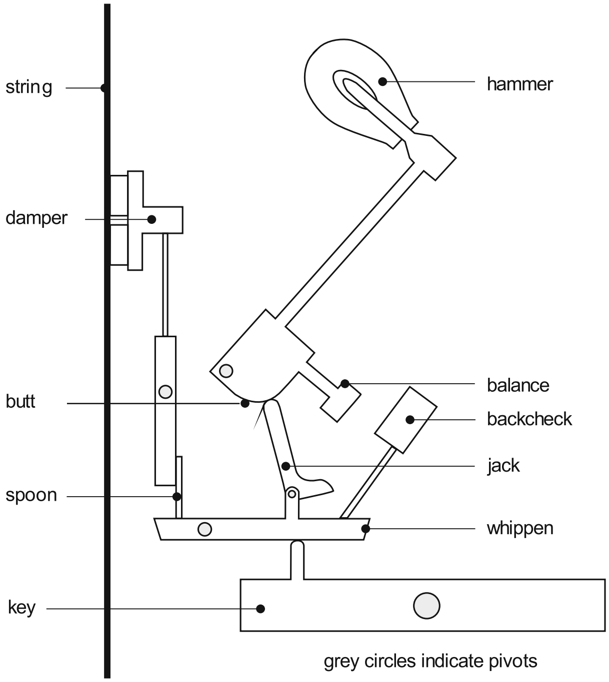


**Fig. 2.** Traditional piano mechanism. A conventional (Comprehensive) animation presents all these components together and faithfully depicts their dynamics (operational details in [3]).

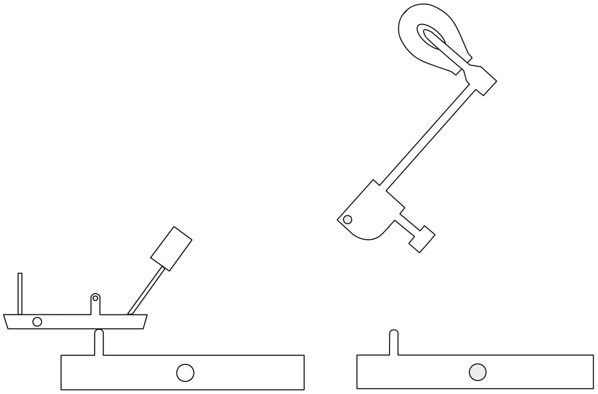


**Fig. 3.** Example Contiguous (left) and Non-contiguous (right) partial animation frames.

were required to learn how the mechanism works, including the ways in which its various components contribute to the piano's overall operation. Their learning was evaluated by assessing the quality of the mental model they developed for the mechanism's operation as a result of studying the animation.

Participants who studied the Contiguous version performed significantly better on a test of mental model quality than those in the Comprehensive and Non-contiguous conditions (Contiguous mental model scores were more than half as large again as scores from the other two conditions) [3]. Eye tracking data were collected using Areas of Interest (AOIs) based on parts of the piano mechanism such that the outline of each AOI captured the region swept out by that component or sub-component during the course of its movement. Further (much smaller) AOIs were located on the regions where two components interacted due to a point of contact. Figure 4 is a notional depiction that illustrates the principle used to define these component and contact AOIs (but in practice, the boundaries of AOIs were extended somewhat to ensure that all relevant fixation data are included). Viewers' foveal fixations on different parts of the display were analyzed in terms of their frequency and duration.

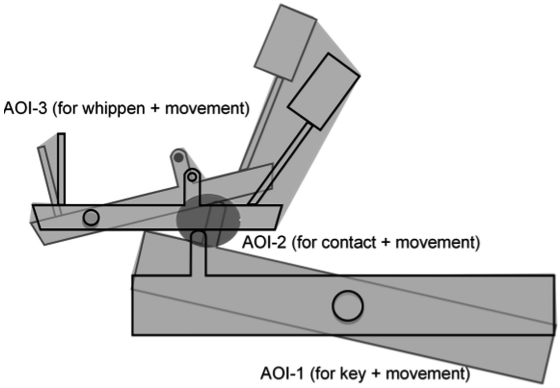


**Fig. 4.** Grey regions provide a notional illustration of the three AOIs used to capture regions swept out by the key, whippen and key riser contact during operation of the piano mechanism.

According to the eye tracking literature, longer total fixation durations invariably occur because more fixations are being made within areas of interest. However, eye tracking data from our experiment revealed an unusual *negative* correlation ($r = -0.81$) between number of fixations made and fixation duration. These fewer, longer fixations were interpreted as an indication of deeper processing. Comparison of the eye tracking videos from the three conditions suggested that participants in the Contiguous condition made particularly prolonged fixations in the small regions of contact between pairs of components (e.g., where the key riser contacts the underside of the whippen, see Fig. 4, AOI-2).

Statistical analysis of the eye tracking data showed that the total fixation duration on each of these contact regions (key-whippen, whippen-jack, jack-butt, etc. etc.) was significantly greater for Contiguous participants (Table 1a). This contact region is of particular importance with regard to the dynamic relations that occur for component pairs presented in the Contiguous condition because it is the crucial link between a cause and its effect.

**Table 1.** Part a of this table presents means (SD) and one way ANOVAs for eye fixation lengths (in seconds) and counts across the total animation exposure time. It reports results from all three conditions (Contiguous, Non-contiguous and Comprehensive, each one N = 20) for (i) contact AOIs in the animation, and (ii) the remaining non-AOI area. Part b of the table concerns just a short subsection of the total exposure time for the Contiguous condition only and compares the foveal processing intensity indices (see explanation below) for the contact and component AOIs.

| Gaze measures | Locations | Contiguous | Non-contiguous | Comprehensive | ANOVAs |
|---|---|---|---|---|---|
| **Part a**<br>*Total animation exposure:4 min*<br>Fixation length | All contact AOIs | 120.02<br>(18.38) | 89.73<br>(18.55) | 103.62<br>(20.67) | $F(2,57) = 12.49$<br>$p < .0001$<br>$\eta p^2 = .30$ |
| | Non AOI area | 107.79<br>(14.35) | 133.62<br>(14.61) | 124.50<br>(19.91) | |
| | Total | 227.81<br>(15.53) | 223.36<br>(22.52) | 228.13<br>(20.51) | |
| Fixation count | All contact AOIs | 204.20<br>(45.03) | 196.15<br>(44.23) | 215.40<br>(61.89) | $F(2,57) = 0.71$<br>$P = .49$<br>$\eta p^2 = .02$ |
| | Non AOI area | 272.65<br>(71.03) | 364.15<br>(57.58) | 327.85<br>(51.00) | |
| | Total | 476.85<br>(107.93) | 560.30<br>(89.39) | 543.25<br>(96.92) | |
| **Part b**<br>*20 s subsection of animation exposure:*<br>foveal processing intensity index (seconds of fixation per square cm) | Contact AOIs | 0.52 | | | |
| | Component AOIs | 0.04 | | | |

Unusual clusterings of fixation activity in the immediate vicinity of the contact point where two components interacted were observed in the Contiguous condition videos taken by the eye tracker. Such clustering was not present in videos from the other two conditions. This observation prompted us to single out the set of eye tracking data obtained from those in the Contiguous condition for more in-depth analysis. Of particular interest were activities concerned with the processing of fundamental cause-effect relationships that are central to APM processing phases 2 and 3. According to the APM, the cause-effect relationships a viewer establishes during these phases are domain-general (rather than domain-specific). In order to establish such a relationship between two interacting components of the piano mechanism, the viewer must characterize not only the contact interaction that links the cause and effect components involved, but also the respective dynamics of those two components (refer to Fig. 4).

We therefore expected further analysis of the Contiguous participants' eye tracking data to show that their concentration of foveal processing on this contact interaction region was accompanied by a similarly close monitoring of component movements. If both of these substantial monitoring tasks were being carried out by foveal processing, fixation activity should be appropriately distributed between the contact AOIs and the component AOIs. To test this expectation, we devised an *index of foveal processing intensity* that expressed fixation durations per unit area (calculated as the ratio between the total foveal fixation duration in each AOI (seconds) and the total area of that

AOI ($cm^2$). Fine grained analysis using this index was concentrated on a 20 s segment of the Contiguous animation comprising the movements of the key, whippen and jack. The AOIs on which this analysis was based captured two classes of eye tracking data (i) fixations on the small areas of contact between adjacent components where cause-effect interactions took place (for key-whippen, whippen-jack and jack-butt) and (ii) fixations on the much larger areas swept out by whole components or major subcomponents as they performed their operational movements (the key, whippen, jack, and butt). Results for the index of foveal processing intensity are shown in Table 1b. Contrary to our expectations, these results indicated that although participants in the Contiguous condition applied a very high level of foveal processing intensity to the regions of contact interaction, they applied a very low level of such processing activity to the components themselves (Table 1b). Further, almost none of the 20 participants ever fixated on the key or the whippen whereas they all made multiple and prolonged fixations within contact AOIs.

This puzzling apparent neglect of information that is absolutely central to establishing causal relationships left us with the question of how participants could be characterizing the respective behaviors of the cause and effect components, if not via foveal processing. In the next section, we suggest an alternative means by which those in the Contiguous condition may have extracted this vital information. To prepare the ground for explaining our suggestion, we first use a specific example to elucidate the types of dynamics involved when two piano components interact.

## 2 Partitioned Perceptual Processing

Figure 5 depicts a subset of the piano mechanism that consists of just the key and the whippen. These two components are in contact at the point where the key's riser meets the whippen's lower surface. Before a piano player depresses the key, this point of contact is located at position $C_1$. Then when the key is played, it rotates clockwise around its pivot and the riser protruding from its top surface pushes on the whippen. This interaction causes the whippen to rotate anticlockwise around its pivot until the

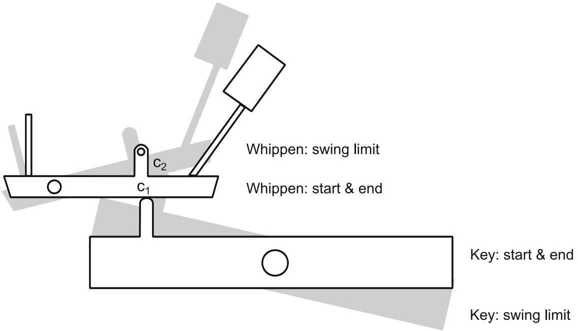


**Fig. 5.** Example showing (i) global cause and effect movements of a pair of piano components and (ii) the local contact interaction that links these together into a causal relationship.

key and whippen both reach the limit of their respective swings. By that moment, the key-whippen contact has reached position $C_2$. Then upon release of the key, its riser retraces its journey back to $C_1$ along the undersurface of the whippen as these components return to their starting positions.

Two very different types of movements are present in this complete cycle. On one hand, there are the macro-scale reciprocal swings of the key (clockwise) and the whippen (anticlockwise). This movement patterns exemplifies a very common type of see-saw behavior exhibited by many everyday devices involving simple levers consisting of a rigid bar that is free to rotate about a pivot. Our extensive experience with such devices equips us with well-developed domain general background knowledge about their typical behavior. On the other hand, the riser-whippen contact interaction that takes place to-and-fro along the path between $C_1$ and $C_2$ occurs on a far more restricted (micro) scale. The exhibited behavior is also highly specific to this particular case of the piano mechanism and therefore not underpinned by the type of domain general knowledge that is available for the macro-scale reciprocal swings of the key and whippen.

We suggest that the extreme differences in these two aspects of the dynamics likely have important implications for viewers' allocation of perceptual resources when they process such pairs of interacting components. Our assumption is that in order to comprehend the role that this type of interaction plays in the piano mechanism's overall functionality, viewers need to be able to comprehensively characterize the cause-effect linkage involved. This requires them to relate (i) the micro-scale details of the continuous contact interaction along the $C_1$ - $C_2$ pathway to (ii) the macro-scale reciprocal swings of the key and whippen that occur during this interaction. A parsimonious way to perceive the information required for establishing this relationship internally would be to process these two aspects of the dynamics in parallel.

Human visual perception in general relies on the complementary, coordinated operation of foveal and peripheral processing [4]. Perception can be optimized by appropriately partitioning these perceptual resources between the various aspects of a set of visual information that confronts the viewer. With regard to the present key-whippen example, we contend that foveal processing should be better suited to the more detailed analysis required for characterizing micro-level information about the riser-whippen contact interactions, while peripheral processing should be better suited to dealing with the macro-level information concerning the overall reciprocal motion of the key and whippen. Such matching of processing type to processing task can be thought of as a form of perceptual partitioning. If this type of partitioning does indeed occur, eye tracking data should indicate that viewers tend to allocate most of their foveal processing resources to the small region in which contact interactions occur, while leaving peripheral processing to take care of the more global movements of the key and whippen.

## 3 Elaborating the APM and Improving Effectiveness

The preceding discussion raises the possibility that peripheral processing can play a far more central role in comprehension of animated diagrams than previously acknowledged. Our empirical results reported above support this possibility. The almost total neglect of peripheral processing in research on learning from animation can perhaps be attributed to influences such as the dominance of eye tracking approaches in this field (a technique based on foveal processing only) and the widespread view that peripheral processing is intrinsically ‘inferior’ to its foveal counterpart, a notion that has been contradicted by recent research. [4]. It may well be time to confront these influences and redress the limiting effects they could have on future progress of animation research. This would require researchers to no longer ignore the possibility that information acquired via peripheral vision may make a substantial and ongoing contribution to animation processing (c.f. [5]).

However, it is important to note that the phenomenon of perceptual partitioning reported in this paper came to light under the very particular circumstances that existed in the Contiguous animation condition (which produced the best mental models). Those circumstances allowed Contiguous participants to devote their foveal processing capacity almost exclusively to the demanding task of analyzing and characterizing details of a contact interaction linking cause to effect. In parallel with this all-consuming foveal activity, they were also able to monitor associated changes of the cause and effect components by delegating such monitoring to peripheral processing. A far less satisfactory alternative scenario for Contiguous participants based on foveal processing alone would have been for them to make multiple fixation switches between highly localized contact interaction dynamics and more global component dynamics (c.f. [6]). In addition to being a much less parsimonious use of processing resources, such an alternative would carry the risk of various disruptions involved in switching between different sites of activity.

Despite not having previously been reported in the animation research literature, such a partitioning between foveal and peripheral processing that enables them to operate in parallel is in fact a normal (rather than exceptional) feature of everyday vision [4], especially in dynamic situations. For example, in situations such as car driving, an individual can automatically monitor a vehicle’s wider dynamic surrounds while at the same time performing detailed, analytical visual interrogation of far more localized information [7]. However, this efficient and highly successful form of resource allocation can of course be severely compromised by misallocation of foveal processing (such as viewing a mobile phone screen during driving) and the serious disruptions that switching between different visual targets typically involves.

The ‘ideal’ processing situation that appeared to pertain for participants who studied the Contiguous version could also be similarly compromised by introducing changes likely to degrade peripheral processing. One way to introduce such change would be to add more information to the surroundings of each of the component pairs used in the Contiguous condition (for example, by including more of the piano’s mechanism in these partial animations). Addition of such ‘clutter’ introduces visual crowding that can have a negative effect on the scope, accuracy and coherence of information extraction via peripheral processing [4]. This may partly explain why the results obtained by animation researchers (who almost exclusively use cluttered conventional comprehensive animations in their investigations) have not alerted them to the possible role of peripheral processing that has been indicated by our present findings.

Although our argument about perceptual partitioning is based on empirical results from only a small subsystem of the piano mechanism (specifically, the key-whippen pair), it seems likely that this phenomenon could apply far more generally. However, instantiation of this visual ‘division of labour’ would require the presentation circumstances to be sufficiently similar to those that characterized the Contiguous animation approach. Key requirements would be a minimal number of components (e.g., a pair), components whose respective movements are related by a contact interaction, and gross component motions that are easily recognized as resembling the familiar dynamics of everyday experience. These circumstances in fact apply throughout the piano mechanism so that perceptual partitioning would be expected to occur for other pairs of components where contact interactions are present (such as the jack and the hammer). Work in progress supports this expectation. Further, we see no reason why this phenomenon should be restricted to the piano mechanism alone. Provided the requirements mentioned above are fulfilled and the subject matter is appropriate, the same type of processing economies could be invoked for many different types of mechanisms.

### 3.1 Elaborating APM Stages 2 and 3

While acknowledging the caveats given above, it seems prudent to review several aspects of the APM (as outlined in its original exposition) to take account of the findings presented in this paper. In the initial version of the APM, it was suggested that once the continuous flux of a presented animation had been decomposed into individual event units during Phase 1, Phase 2 processing could proceed during which the viewer connects two *or more* event units at a time into superordinate composite structures that we termed ‘dynamic micro chunks’. However, if a viewer is able to invoke the form of partitioned perception discussed in this paper, it would presumably be best to process event units in pairs, rather than in larger groups. Pairwise processing should allow a highly efficient allocation of perceptual resources in which there is a near optimal match between (i) the processing aspect that is engaged (foveal or peripheral) and (ii) the task to which that aspect is applied (detailed analysis of contact interaction or global characterization of cause-effect dynamics). The efficacy of dealing with dynamic targets in a pairwise fashion has been clearly demonstrated in research with air traffic controllers [8].

With respect to Phase 3 processing, the initial version of the APM is somewhat lacking in processing detail about just how the dynamic micro chunks formed during Phase 2 become connected up by bridging relations to form a superordinate structure of causal chains. A pairwise approach similar to that posited for Phase 2 also seems applicable to Phase 3 since it could be based on the same type of perceptual partitioning. To make this more concrete, consider two possible combinations of event units that could be formed during Phase 2 processing of a conventionally designed (comprehensive) piano animation: (i) a key-whippen dynamic micro chunk and (ii) a jack-hammer micro chunk. The linking up of these two chunks as part of a causal chain is via the pivot that attaches the base of the jack to the riser of the whippen. In terms of the perceptual partitioning approach, the task of characterizing what happens in this highly localized site would be allocated to foveal processing. The detailed analysis occurring in this small region would be complemented by peripheral processing of the dynamics of the proximal whippen (cause) and the jack (effect) rather than the dynamics of the more distant hammer and key (Fig. 6). Note that Fig. 6 is merely a stylized conceptual representation of this possibility; in reality, there would be gradual degradation of the peripheral information with distance from the centre, rather than the sudden change depicted here.

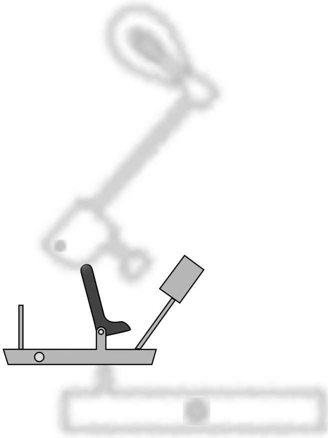


**Fig. 6.** Hypothetical situation in which perceptual partitioning could be invoked (conceptual representation only). Partitioning is confined to peripheral field closest to contact interaction.

The elaborations of APM phases 2 and 3 suggested in the foregoing discussion have been considered only from the perspective of the role that perceptual partitioning could play in the progressive composition of individual event units into domain general causal chains. However, this perceptually-oriented account should be complemented by a consideration of how contributions from top-down processing could modulate the apportioning of perceptual processing discussed earlier. For example, in the case of the piano animation, it is highly improbable that a top-flight piano repair technician (i.e., someone with domain specific expertise in the animation's subject matter) would follow the same processing route during viewing as those who lack specialist knowledge in this domain. Instead, the technician's focus is likely to be on the finer points of piano functioning (rather than its basic operation), with foveal processing used extensively to interrogate these aspects of the mechanism's dynamics.

## 3.2 Intervening to Improve Conventionally Designed Animations

Researchers have considered a variety of factors that may influence comprehension of animations [9], ranging from the dynamic spatial ability of the viewer [10] to the forms of support that are provided to accompany presentation of an animation [11]. Most interventions intended to improve animation processing have, at best, met with limited success. However, it is possible that the phenomenon of perceptual partitioning may provide a more promising basis for devising supportive interventions, not the least because it appears to be theoretically robust and is derived from empirical evidence.

Pronounced partitioning of perceptual resources by which foveal and peripheral processing were optimally allocated to local and global aspects of cause-effect relation formation occurred only for participants who studied the Contiguous version of the animation. However, because the design of this version according to the Composition Approach involves a radical departure from the currently prevailing entrenched design orthodoxies, this strategy for improving effectiveness is something not likely to be widely adopted by animation designers overnight. This raises the question of whether or not it would instead be possible to devise other ways of obtaining the processing efficiencies afforded in the contiguous condition but with a conventional comprehensive animation design rather than one designed according to the Composition Approach.

One possibility could be to approximate the type of situation that exists with a contiguous version pair by applying a suitable intervention to an existing conventionally designed comprehensive animation. If we consider a comprehensive animation of the piano mechanism as an example, perhaps the desired processing affordances could be obtained by 'visually suppressing' all parts of the mechanism except for a pair of event units (e.g., the key and whippen) using anti-cueing techniques such as fading (Fig. 7). Changing the region across which this anti-cueing is applied in a stepwise fashion over time should produce a situation that presents a succession of pair-wise processing opportunities resembling those that were provided in the Contiguous condition. Empirical research is needed to investigate this and other intervention strategies that have the potential to stimulate beneficial perceptual partitioning.

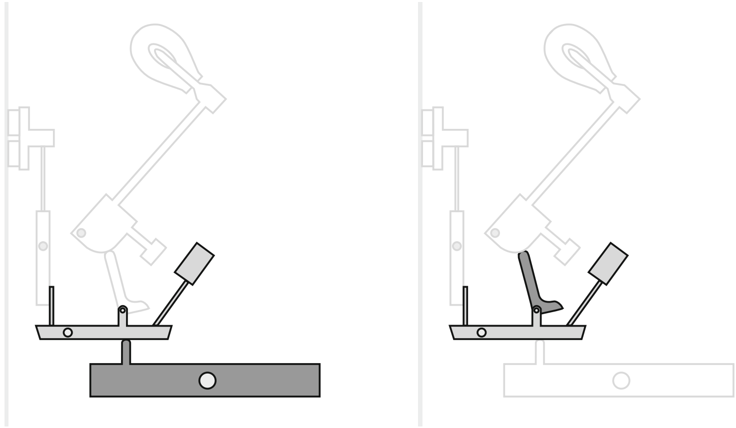


**Fig. 7.** Possible use of anti-cueing to produce a situation resembling that available in a composition approach (i.e., Contiguous pairs) but with a conventionally designed animation.

## 4 Discussion and Conclusion

Our motivation for updating the APM in light of recent empirical and theoretical work is a continuing quest to develop a principled basis for redressing the mismatches between design features and human information processing that exist with conventional animations. Effective extraction from an animated diagram of information about dynamics plays a crucial role in building a high quality mental model of the depicted subject matter because this behaviour indicates the causality that underpins the operation of a system. One feature of research into animation processing to date has been the relative neglect of perception (compared with cognition). This is a considerably more important issue for dynamic displays like animated diagrams than it is for static diagrams because of the powerful influence that dynamics have on perception (and hence information extraction).

We were initially alerted to the possibility of a previously unreported type of perceptual processing by unexpected patterns of fixation in eye tracking videos from a recent experiment. Empirical evidence subsequently gathered from further analysis of the eye tracking data suggested that not only foveal but also peripheral perception can be important in processing animated diagrams efficiently and effectively. For the Contiguous paired presentation, this evidence indicated that, in essence, foveal processing was being devoted exclusively to close monitoring of contact interaction between the components in a pair leaving the task of characterizing the associated overall movement patterns of those components to peripheral processing. More specifically, available perceptual resources were being allocated in parallel according to individual task requirements: detailed analysis of micro-scale dynamics to foveal processing and broad characterization of highly familiar everyday macro-scale dynamics to peripheral vision. Despite the findings being highly novel (and unexpected), this form of tailored resource allocation is not in fact exceptional but rather perfectly normal in everyday visual perception. The findings are also highly consistent with a central aspect of the APM: the composition of event units into more inclusive knowledge structures.

The likely implications for elaborating the APM are that (i) perceptual processing plays crucial role not only in extracting individual event units from an animation's dynamic flux (Phase 1) but also in contributing to the composition of these basic building blocks into higher order information structures such as causal chains (Phases 2 and 3), and that (ii) considerable benefits can be achieved by fostering perceptual partitioning in which responsibility is shared between foveal and peripheral processing resources. However, it appears that such partitioning is contingent on the dynamic subject matter being offered in a suitable fashion (in the case considered here, this was according to the specific pairwise presentation regime available in the Contiguous condition).

The Contiguous animation discussed in this paper was devised primarily for research purposes. Despite its effectiveness, we did not intend it to be adopted 'as-is' by practicing animation designers. Rather, we acknowledge the reality that conventional approaches to designing animations will continue to be dominant into the foreseeable future. However, the situation that exists in unsupported Comprehensive animations is the antithesis of what is required to allow highly efficient partitioning of perceptual resources. For this reason, we are interested not only in using our findings to elaborate the APM, but also in using the insights gained to suggest related interventions (such as the use of anti-cueing to support pairwise processing) that may improve the effectiveness of comprehensive animations. It would also be important to empirically test the effectiveness of using anti-cueing with comprehensive animations (as proposed in this paper) to simulate the type of pairwise processing situation found in the Contiguous condition. The perceptual partitioning finding appears likely to be generalizable to animations of many other mechanical systems that are based on similar types of cause-effect relationships (such as the toilet cistern studied by Hegarty and colleagues [12]). However, this possibility needs to be investigated by future empirical research.

## References

1. Lowe, R., Boucheix, J.-M.: Learning from animated diagrams: how are mental models built? In: Stapleton, G., Howse, J., Lee, J. (eds.) Diagrams 2008. LNCS (LNAI), vol. 5223, pp. 266–281. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87730-1_25
2. Lowe, R., Boucheix, J.-M.: Dynamic diagrams: a composition alternative. In: Cox, P., Plimmer, B., Rodgers, P. (eds.) Diagrams 2012. LNCS (LNAI), vol. 7352, pp. 233–240. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31223-6_24
3. Lowe, R.K., Boucheix, J.M.: Principled animation design improves comprehension of complex dynamics. Learn. Instr. **45**, 72–84 (2016)
4. Rosenholtz, R.: Capabilities and limitations of peripheral vision. Ann. Rev. Vis. Sci. **2**, 437–457 (2016)
5. De Koning, B.B., Tabbers, H.K., Rikers, R.M.J.P., Paas, F.: Animation guidance in learning from a complex animation - seeing is understanding? Learn. Instr. **20**, 111–122 (2010)
6. Bauhoff, V., Huff, M., Schwan, S.: Distance matters - spatial contiguity effects as trade-off between gaze switches and memory load. Appl. Cogn. Psychol. **26**, 863–871 (2012)
7. Strayer, D.L., Drews, F.A., Johnston, W.A.: Cell phone-induced failures of visual attention during simulated driving. J. Exp. Psychol. Appl. **9**(1), 23–32 (2003)
8. Bisseret, A.: Représentation et décision experte: Psychologie cognitive de la decision chez les aiguilleurs du ciel. Octarès Editions, Toulouse (1995)
9. Bernay, S., Bétrancourt, M.: Does animation enhance learning? A meta-analysis. Comput. Educ. **101**, 150–167 (2016)
10. Sanchez, C.A., Wiley, J.: The role of dynamic spatial ability in geosciences text comprehension. Learn. Instr. **31**, 33–45 (2014)
11. de Koning, Björn B., Jarodzka, H.: Attention guidance strategies for supporting learning from dynamic visualizations. In: Lowe, R., Ploetzner, R. (eds.) Learning from Dynamic Visualization, pp. 255–278. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56204-9_11
12. Hegarty, M., Kriz, S., Cate, C.: The roles of mental animations and external animations in understanding mechanical systems. Cogn. Instr. **21**(4), 209–249 (2003)

# Visualizing Conversational Structure: Effects of Conversation-Analytical Knowledge and Social Media Experience

Leonie Bosveld-de Smet[1](✉) and Anne-Roos Bakker[2]

[1] Department of Information Science,
University of Groningen, Groningen, Netherlands
`l.m.bosveld@rug.nl`
[2] Embrace SBS, Groningen, The Netherlands
`anne-roos.bakker@embracesbs.com`

**Abstract.** This paper explores how people understand and visualize externally a synchronous multi-threaded four-party conversation that was audio-recorded, and investigates whether conversation-analytical knowledge and/or digital skills with social media tools have an effect on the nature and complexity of the conversational structure depicted in the representation constructed. An experiment has been performed, in which 60 participants took part. Their task was to listen to a conversation, and to display it on a magnetic whiteboard in their own way. Predesigned conversation's utterances and pictures of participants were provided, as well as markers of different colors. Both visualization process and product were coded. Coding of process included production time and relistening behavior. The product was analyzed with respect to the ordering of utterances. We used four characterizations of ordering: by chronology, by reply-to relationships, by topic, and by conversational participant. Production time and relistening behavior turn out to have varying effects on products. Results of the representations' analysis suggest that conversation-analytical knowledge or experience with a variety of social media influence the type and the number of ordering principles used.

**Keywords:** External representation · Conversational structure
Conversation analysis · Digital skills

## 1 Introduction

In visual communication, application domains are linked to graphical domains. An interpretational link maps application domain entities to graphical ones. Application domain entities are inherently visible or not. Visibility may influence the graphical marks chosen by people to construct external representations [4]. Especially in the invisible, abstract case it is interesting to see how people conceptualize the abstract information in the application domain and represent it externally. As pointed out by [1], self-constructed external representations are affected by individual differences, e.g. prior knowledge. [3] shows that abstract

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 484–491, 2018.
https://doi.org/10.1007/978-3-319-91376-6_43

knowledge is involved in the application of diagrams visualizing different informational structures. Both the visualization of something in the world that is not inherently visible, and the influence of more or less abstract domain knowledge, will be addressed in this paper. The abstract entity involved is a multi-party, multi-threaded conversation, and in particular the underlying structures it can be associated with. Different groups are targeted in this study. On the one hand we distinguish people with and without some prior knowledge regarding abstract conversational notions, such as turn-taking, mentioning and topic shift. On the other hand we focus on differences in computer-mediated communication experience with a variety of social media tools, such as Facebook, Reddit and text chat. The outcomes of this paper will give insight in (i) external representations of conversational structure, and (ii) effects, on these representations, of prior knowledge, either of abstract conversational notions or of a broad repertoire of different online discussion representations. This study is part of a larger research where we investigate visualizations of conversations, not only from a production, but also from a comprehension perspective. Results may provide clues for optimizing user interfaces of applications representing group discussions.

This study focuses on conversational structure. A conversation is defined as a collection of related utterances, also referred to as messages. The utterances are bound by various relationships, based on different aspects involved in conversations. One such aspect is chronology, which corresponds to a sequential ordering of utterances by time. Interactional coherence is another aspect, which is based on reply-to relationships between utterances. If emphasis is put, not on chronology, but on coherence, conversations can be viewed as having a tree structure. Much more ordering approaches are imaginable [2].

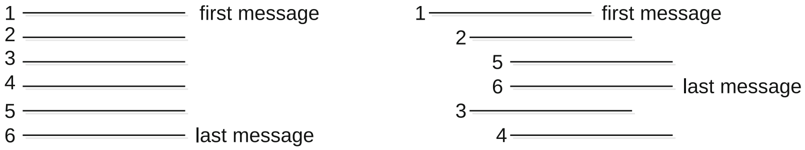


**Fig. 1.** Message orderings according to the sequential (left) and tree (right) model.

One branch of research interested in the analysis of computer-mediated conversations concerns the effective and efficient visualization of online discussions. As pointed out in [5], implicit in online conversational designs are two models of conversation, the sequential and the tree model. Figure 1 shows schematic representations of these models. A visualization supporting the sequential model allows to answer questions about the chronological order of utterances (e.g. 'which of these two messages was sent first?') at a glance, that is, with direct simple visual inspection. Text chat interfaces are often based on this model. A visualization based on the tree model allows to answer questions about reply-to relationships between utterances (e.g. 'Does this message have any replies?') at a glance. Reddit's interface supports the tree model. Facebook and many online discussion sites use a mixed model, in which, however, the tree model is generally not applied in a consistent way; comments are presented as responses to an original post, but are not ordered with respect to reply relations among one another. We have added two other ordering criteria of utterances to the ones discussed above, which we will refer to as the topic and the participation model. Topic clusters utterances based on the subject discussed. Participation aggregates messages according to their contributor. Visualizations supporting the topic model allow to answer questions such as ‘Which topic does this message address?’ at a glance, ones based on participation facilitate answering questions such as ‘Who contributes to the discussion with this message?’.

## 2 Method

This study aims to address the following research questions: RQ1: Which model types (sequential, tree, topic, or participation) and which model tuples (combinations of model types) is/are supported by the visualizations individuals produce of the conversational structure, implied in a multi-party and multi-threaded conversation? RQ2: Does more or less abstract knowledge of conversation have an effect on the model type and tuple used in the visualizations produced? An empirical experiment has been conducted in order to answer RQ1 and RQ2.

### 2.1 Participants: Recruitment and Description

Sixty persons, aged between 22 and 50 years, were recruited for participation in this experiment. We have made a distinction between two types of prior knowledge, conversation-analytical knowledge, and knowledge based on experience with computer-mediated conversations via a variety of interfaces. The first one implies abstract conversational concepts, rules and patterns, either acquired as communication specialists or as graphical designers of online conversations. The second type is based on concrete examples and is acquired by reading and/or participating in online conversations in various applications, viz. Facebook, WhatsApp, Reddit, text chat, and online discussion boards. Thirty participants (14 women and 16 men) with conversation-analytical knowledge were recruited. The other half (15 women and 15 men) was selected as not having prior abstract knowledge. A questionnaire at the end of the experiment was used to get information about the experience of the participants with online conversations. Analysis of the questionnaire answers shows that all 60 participants use Facebook and WhatsApp regularly. Remarkably, participants who indicated to use Reddit, turn out to participate more in chat conversations and online discussions. We decided to consider this group (N = 27) as the participants with a broad experience with

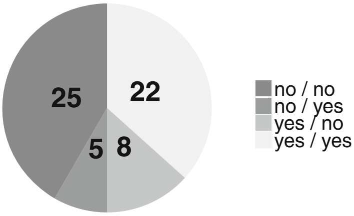


**Fig. 2.** Distribution of participants (N = 60) with respect to conversation-analytical knowledge and/or experience with social media tools.

different social media tools. As shown in Fig. 2, the majority of the preselected knowledgeable participants (yes/no and yes/yes segments) turns out to be experienced (yes/yes). We think that this relation, which is significant ($\chi^2(1, \mathrm{N} = 60) = 19.46$), is not causal, but just a coincidence. A few non-knowledgeable participants are experienced (no/yes segment). What does not seem to be fortuitous is the predominance of males in the group of experienced participants (8 women versus 19 men).

### 2.2 Design and Procedure

The experiment has a between-group factorial design. All participants were individually exposed to the same experimental conditions. They got the task 1) to listen to two conversations, a simple and a more complex one, between four conversational participants (the same for both conversations: two women and two men, with distinguishable voices), and 2) to visualize them on a whiteboard. The simple conversation (of 14 s) consisted of five utterances continuing the same topic, and served as a means to get the participants acquainted to the type of experiment. The second, more complex, conversation (of 49 s), which is the focus of this study, consisted of 24 utterances, and involved two main different topics, each containing subtopics; the four conversational participants discussed the buying of a present for a mutual friend (topic 1), and the cooking of a meal (topic 2). Discussion of one (sub)topic was disrupted by discussion of other (sub)topics. Relistening was allowed. We chose to expose the participants to an audio-recording of these conversations and to allow relistening in order to reduce the cognitive load. The exact wording of the task was (translated from Dutch): 'Please represent the conversation you hear on the whiteboard in your own way. There is no time limit. You may relisten the conversation, or parts of it. You are allowed but not obliged to use the materials provided.' With this succinct wording, we hoped to elicit as spontaneous productions as possible within the boundaries of the experiment (see materials below). The wording has been kept constant for each participant. Several aspects of the visualization process were recorded. Photographs were taken of all visualization products.

### 2.3 Materials and Coding

The experimental materials consisted of a magnetic whiteboard, small white rectangular magnets corresponding to the conversation's utterances (five for the first and 24 for the second conversation). The utterance text (in black) preceded by name of utterer (different color for each of the four utterers) were printed on the magnet. Four circular magnets with photo and name of each conversational participant were available, as well as markers of different colors, and a wiper. We have chosen to provide these aids in order to relieve the participants and to steer toward listening carefully to and laying out the whole conversation.

The construction process was coded with respect to construction time and number of times that (part of) the conversation was relistened. Coding of the final visualization product was performed by two coders and based on the four

**Table 1.** Frequencies of model types supported by the visualizations constructed (N = 104).

| Model Type | | | | Total |
|---|---|---|---|---|
| C (%) | R (%) | T (%) | P (%) | |
| 51 (49) | 9 (9) | 27 (26) | 17 (16) | 104 (100%) |

model types introduced in the introduction. Different series of questions were answered in order to decide which model type(s) is/are supported by the representation laid out on the whiteboard. The following labels were used to characterize the conversational structure visualized: C (chronology) , R (reply-to), T (topic), and P (participation), for the sequential, tree, topic and participation model respectively. Theoretically, each ordering criterion (C, R, T, P) and its combination with other criteria (e.g. CR, CT, CTR, CRTP) could end up as a visualization label. There was a high degree of agreement (96%) about the final labels for each product.

# 3 Results

## 3.1 Visualization Process and Products

Every participant was able to depict the conversation on the whiteboard properly, that is, design a visualization corresponding correctly to aspects of the conversation overheard. In most representations (93%), all 24 utterances are used. Only 25 participants (42%) decided to use all discussants' profile pictures, and more than half (58%) chose to augment their representations with their own graphical marks (mostly lines or arrows), using the color markers. There is quite some variation in the time spent for relistening and construction. The mean time is 8.1 min (SD = 4.3; mode and median = 7; range = [3–25]).

Three (5%) participants did not relisten (parts of) the conversation, 18 (30%) did only once, and 38 (63%) (one missing value) more than once. Relistening the conversation more than once tends to take more time to produce the visualization than relistening zero times or only once. The relation approaches significance (t(57) = 1.9547, p = 0.0556).

Tables 1 and 2 show the results of the visualizations produced.

**Table 2.** Frequencies of (combinations of) models supported by the visualizations constructed (N = 60).

| Single | | | | Double | | | | | | Triple | | | | Quadruple | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C | R | T | P | CR | CT | CP | RT | RP | TP | CRT | CRP | CTP | RTP | CRTP | |
| 16 | 0 | 4 | 4 | 1 | 15 | 11 | 1 | 0 | 0 | 6 | 1 | 1 | 0 | 0 | 60 |
| 24 (40%) | | | | 28 (47%) | | | | | | 8 (13%) | | | | 0 (0%) | 60 |

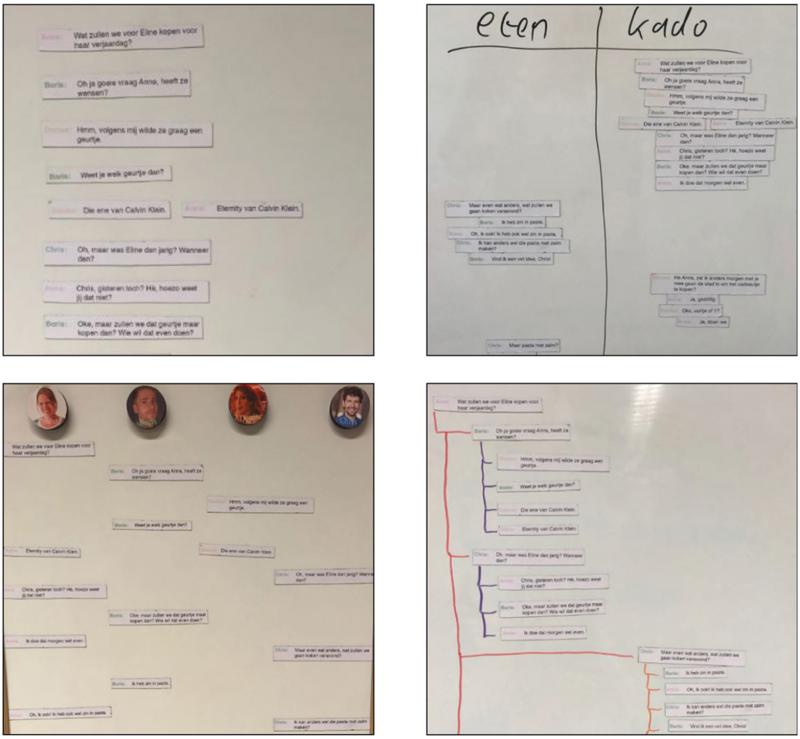


**Fig. 3.** Illustrations of the visualizations supporting the C, CT, CP and CRT models.

On average, each participant bases his/her visualization of the conversation on 1.73 models. All models are used, but in different proportions. The sequential model prevails, the tree model is applied the least, and the topic model is used more frequently than the participation model (see Table 1). The difference in proportion is significant at the 0.01 level ($\chi^2(3, \mathrm{N} = 104) = 38.4$).

Table 2 shows visualizations supporting the quadruple don't occur, while single and double models occur more often than triple models. When we oppose single models (40%) to mixed (double and triple) ones (60%), the difference in proportions is not significant ($\chi^2(1, \mathrm{N} = 60) = 2.4$), but it is when we oppose single and double models to triple ones ($\chi^2(2, \mathrm{N} = 60) = 9.8$). There is no relation between relistening behavior and complexity of visualization, that is, the tuple based on, nor between relistening and type of model used to layout the conversation's utterances.

In general, we see clear patterns in the diversity of the external representations participants produce. The models most frequently supported are the C-, CT-, CP-, and CRT-model (see Table 2). Each picture in Fig. 3 illustrates visualizations of each of these models.

**Table 3.** Frequencies of model types supported by the visualizations constructed (N = 104), per knowledge group.

| Type of prior knowledge | Model Type | | | | Total |
|---|---|---|---|---|---|
| | C (%) | R (%) | T (%) | P (%) | |
| yes - yes (N = 22) | 18 (35) | 5 (56) | 17 (63) | 4 (23) | 44 (42%) |
| yes - no (N = 8) | 8 (16) | 2 (22) | 3 (11) | 1 (6) | 14 (13%) |
| no - yes (N = 5) | 5 (10) | 0 (0) | 2 (7) | 3 (18) | 10 (10%) |
| no - no (N = 25) | 20 (39) | 2 (22) | 5 (19) | 9 (53) | 36 (34%) |
| | 51 (49) | 9 (9) | 27 (26) | 17 (16) | 104 (100%) |

### 3.2 RQ2: Effects of Prior Knowledge on Model Type and Tuple

Conversation-analytical knowledge has an influence on the model type(s) chosen for the construction of the conversational structure. While use of the sequential model is equally distributed among participants, the topic model shows up more often in the visualizations of the knowledgeable group, and the participation model occurs more in those of the non-knowledgeable group (see Table 3). The proportional differences between topic and participation is significant ($\chi^2$(1, N = 44) = 8.6). The distributional frequencies between the two groups of participants differentiated by experience demonstrate a similar pattern, but without significant differences ($\chi^2$(1, N = 44) = 3.57). In general the tree model is applied sparsely, but more frequently by the knowledgeable group than by the non-knowledgeable one. Remarkably, there is no clear preference for the tree model by experienced participants, who are acquainted with Reddit's tree design of discussions.

Figure 4 shows the use of simple and more complex models by the different groups of participants. Participants with conversation-analytical knowledge base their constructions significantly more often on a mixed (double or triple) model than on a single one ($\chi^2$(1, N = 60) = 6.94). And so do experienced participants ($\chi^2$(1, N = 60) = 9.44).

Although the experienced group contained more men than women, no gender effect has been observed; neither type nor complexity outcomes are influenced by gender.

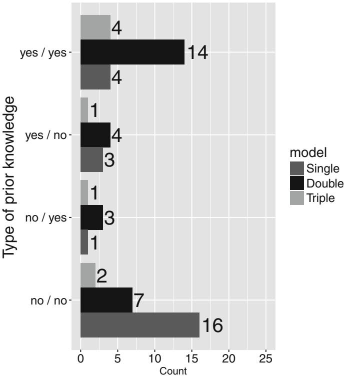


**Fig. 4.** Frequencies of model tuples, per group.

For all groups, the relation between the model complexity and the production times follows the same pattern. Applying more complexity takes more time. An exception is the group with no conversation-analytical knowledge, but with experience. This may be due to the small numbers involved (no-yes: N = 5). No significant relations are observed between knowledge/experience groups and relistening behavior.

## 4 Discussion

The sequential model is supported by almost all visualizations constructed. This does not come as a surprise, given the stimulus, which is a synchronous discussion, in which participants apply turn-taking rules properly. The poor use of the tree model, even by Reddit users, may be explained by the effort it takes to detect reply-to relations in the conversation overheard. Conversation-analytically knowledgeable participants base their constructions more on the topic model, while the participant model is predominant in those of the group without abstract knowledge. The availability of participants' pictures in combination with the relatively easy attribution of utterances to each conversational participant probably has steered the non-knowledgeable group to use the participation model rather than the topic model. Identifying different conversational topics and relating them to utterances requires more conversation-analytic knowledge than identifying conversation participants and cluster their utterances.

Both the conversation-analytical knowledgeable group and the group with social media experience create more complex visualizations of conversational structure, without taking more time or relistening more often, than the non-knowledgeable and the non-experienced group. We can conclude that knowledge, either of conversational abstract notions or of a broad repertoire of online discussion interfaces, allow people to recognize and identify more conversational aspects, and to give indications of how to represent them externally. Conversation-analytical knowledge, or experience, or both have an effect, and future research should give insight in the exact nature of prior knowledge.

The quadruple model was not supported by any visualization. This may have several reasons. It may be cognitively too demanding to remember more than three utterance-related aspects of conversations overheard, the depiction of four aspects risks to give rise to cluttered depictions, or participants just don't get the idea that this is an option as well. Existing interfaces of commonly used social media tools don't give any clues how to visualize more than three orderings. This finding also gives rise to future research.

## References

1. Cox, R.: Representation construction, externalised cognition and individual differences. Learn. Instr. **9**(4), 343–363 (1999)
2. Donath, J., Karahalios, K., Viegas, F.: Visualizing conversation. J. Comput.-Mediat. Commun. **4**(4) (1999)
3. Hurley, S.M., Novick, L.R.: Context and structure: the nature of students' knowledge about three spatial diagram representations. Think. Reason. **12**(3), 281–308 (2006)
4. Tversky, B.: Visualizing thought. Top. Cognit. Sci. **3**(3), 499–535 (2011)
5. Venolia, G.D., Neustaedter, C.: Understanding sequence and reply relationships within email conversations: a mixed-model visualization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 361–368. ACM (2003)

# Comparing and Contrasting Within Diagrams: An Effective Study Strategy

Jennifer G. Cromley[1](✉) and Kyle Mara[2]

[1] University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA
jcromley@illinois.edu
[2] University of Southern Indiana, Evansville, IN 47712, USA

**Abstract.** The study strategy of comparing-and-contrasting has been well validated for learning from text, but not from diagrams. As part of a semester-long study strategies intervention for undergraduate biology students, we created 4 short instructional videos demonstrating the strategy of comparing-and-contrasting within diagrams (CC DIA) and delivered these just before the first course exam. We hypothesized that this strategy would help students develop a deeper comprehension of the instructed biology content. Participants were 128 undergraduates in a 2nd semester introductory (molecular and cellular) biology course, who participated in exchange for extra course credit. Students who accessed our videos scored a significant 5.5% points higher on the first exam of the semester, compared to students in other conditions or non-viewers ($d$ = .35). Our brief (approx. 10 min per week × 4 weeks) instruction in using diagrams to learn biology yielded significant gains in undergraduate achievement.

**Keywords:** Strategy instruction · Compare-and-contrast · Biology

## 1 Introduction

### 1.1 Comparing and Contrasting as a Study Strategy

The study strategy of comparing-and-contrasting has been well validated for learning from text [1], but not as much from diagrams [but see 2]. When students engage in the compare-and-contrast strategy, they use two stimuli about similar topics (e.g., 2 texts or paragraphs, one about mitosis and one about meiosis; 2 diagrams, one about animal cells and one about plant cells) and articulate what is similar across the two and what is different across the two. Comparing-and-contrasting (CC) is cognitively complex, in that it draws on numerous cognitive processes. CC draws on analogical reasoning ("this is like that"), which necessarily involves metacognitively monitoring the state of one's knowledge [3]. It also draws on deductive reasoning, such as reasoning through why two historical figures wrote differently about the same event [4]. CC may also involve elaborative inferences that incorporate prior knowledge with information from a text, such as knowledge about colonialism in Central America to inform learning from multiple texts on the Panama Canal [5]. In addition, CC in diagrams usually involves matching identical or analogous structures, which are sometimes indicated by identical color codes, arrows, or other conventional diagram features. Therefore, CC in diagrams

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 492–499, 2018.
https://doi.org/10.1007/978-3-319-91376-6_44

likely requires a better understanding of conventions of diagrams than does interpreting a single image within a diagram.

A small body of research has examined comparing and contrasting with diagrams and other visualizations. For example, in an influential 1998 paper, Bransford and Schwartz [6] reported on a highly effective teaching method they called “contrasting cases,” in which students are given multiple examples to learn from simultaneously. In some cases, the examples included realistic drawings, photographs, schematic diagrams, and graphs. Comparing-and-contrasting among these visualizations led to better conceptual learning of content; two contrasting cases presented simultaneously were more effective than multiple cases presented simultaneously.

In a meta-analysis of learning with contrasting cases, Alfieri et al. [7] found that cases targeting perceptual skills (e.g., find common elements across a set of drawings, compare and contrast across videos) had even larger effects on learning (*d* = .72) than those focused on declarative or procedural skills (*d* = .54 and .40, respectively). More recently, Schunn and colleagues (under review) delivered contrasting cases instruction in hundreds of middle school classes, sometimes using textual contrasts and sometimes diagrams. They found that CC students performed significantly better than students not receiving CC instruction on both researcher-developed diagram comprehension questions and on diagram questions from statewide end-of-course standardized tests.

These large effect sizes from instruction suggest that students are not already using CC to learn, and teachers are not routinely providing effective instruction in CC. For example, Cromley et al. [8] found few examples of comparing and contrasting when undergraduate biology students—not provided with any training about diagrams—provided think-aloud protocols from their own course textbook (see Fig. 1), even when the diagrams invited such comparisons.

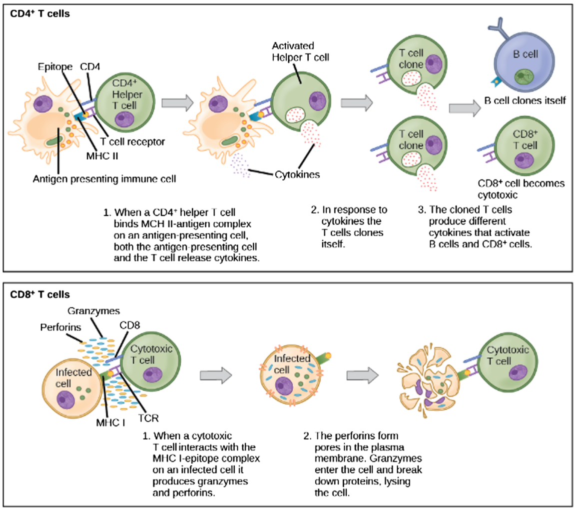


**Fig. 1.** A diagram analogous to one used in used in a think-aloud study by Cromley et al. [8]. Open-source image from Wikipeda.

In addition, the nature of biology learning suggests that contrasting cases should be a particularly effective instructional technique. A major theme in biology is the relationship between biological structure and biological function [9]. That is, the particular biological elements and their inter-relationships (e.g., alveoli, bronchi, ribs and the muscles around the lung; stomata, veins, epidermis, palisade cells, and mesophyll in leaves) are directly related to the functions of those parts (intake of oxygen and discharge of carbon dioxide in the lungs; water intake for photosynthesis in leaves). Teaching diagram comprehension with single representations may lead students to focus too much on structure, whereas contrasting diagrams may invite more considerations of function—Why do these structures (e.g., in plant and animal cells) differ? *Because they must perform different functions.* Taken together, the literature review and our understanding of the structure-function challenge in learning science suggests that CC in diagrams is a promising instructional technique for science learning, but one that has rarely been systematically tested at the undergraduate level.

We therefore developed a series of instructional materials, rooted in students' own course, designed to teach them how to compare and contrast within the diagrams in their textbook. To give a sense of what this task involves, consider Fig. 2, where a molecule in chromosomes called histone is shown three times. On the far left, a small number of histone proteins are shown before the histone combines with acetyl groups (small molecular components). This is linked with dashed lines to multiple un-acetylated histones which form "spools" that DNA wraps around. Between part (a) and the left image in part (b), it is critical for the reader to recognize that the same molecule (histone) is shown, it is unacetylated, and that DNA is wrapped tightly around the histone molecules. Between the left part of (b) and the right part of (b), there is a change; the acetyl groups have attached to the "tails" of the histone molecules (labeled in (a)), because of this the DNA unwinds from the histone "spools," and as the caption explains, this unwinding allows transcription to happen. Comparing and contrasting is thus a powerful strategy for making sense of the diagram, and for linking the structures (histone, "tails," DNA, acetyl groups) to functions (unwinding the tightly packed chromosome so that DNA transcription can happen; see another example in Table 1).

We expected that our instruction engaging in comparing-and-contrasting in diagrams would help students better learn the biology content taught in their course. Specifically, we hypothesized that accessing the CC DIA instructional materials from chapters that were tested in the first exam of the semester would lead to higher scores on that first exam, compared to students who did not access the CC DIA instructional materials.

## 2 Method

### 2.1 Participants

Eligible students were enrolled in a 2nd semester undergraduate biology course intended for science majors (e.g., biology, biochemistry, bioengineering, kinesiology), typically taken by Sophomores, at a large, urban, East Coast US research-focused university. Participants were 128 students who had enrolled in the study before the first exam of the semester, were assigned to an experimental condition, and completed Exam 1. They were 64% female and were mostly Sophomores (78%), with 41% from families where neither parent had a Bachelor's degree. They were racially diverse, with 38% self-identifying as White, 27% as East Asian, 14% as South Asian, 7% as Black, and 14% as belonging to other or multiple races.

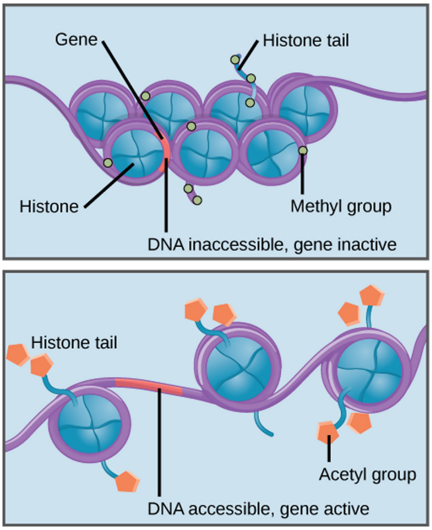
## 2.2 Measure

In this paper, we analyze scores on exam 1, which was a professor-designed exam given during a single 50-min class period in October, 2015. Exam scores had a mean of 75.0 (*SD* = 15.9), ranged from 36 to 108, and were normally distributed.

## 2.3 Intervention Materials

We describe our Compare-and-Contrast in diagrams intervention materials, which were part of a larger semester-long study delivered via the Blackboard learning management system. Students accessing the Compare-and-Contrast in diagrams videos also received one of three motivational supports not described here which were delivered either 1 week before or 1 week after the first exam.

To create our Compare-and-Contrast in diagrams videos, the first author began by identifying diagrams within students' textbook that would be covered in weeks 1–4 of the semester. Among these diagrams, we sought multi-part diagrams where the compare-and-contrast strategy would be required (see Fig. 2).



Methylation of DNA and histones causes nucleosomes to pack tightly together. Transcription factors cannot bind the DNA, and genes are not expressed.

Histone acetylation results in loose packing of nucleosomes. Transcription factors can bind the DNA and genes are expressed.

**Fig. 2.** A diagram analogous to the one used to create the script and subsequent video for the Compare-and-contrast in diagrams strategy instruction. Open-use image from Wikipeda.

After identifying the diagrams, we then drafted a written script demonstrating the strategy within each diagram; this script followed 6 steps in the Pressley and Harris [10] strategy instruction model (see Table 1). Within each script, the strategy was demonstrated 3 times (i.e., a total of 12 diagrams over 4 textbook chapters). After each diagram, we asked the students to "Keep going with comparing and contrasting in diagrams, try this in Figure [X] on page [X]. Pause the video and see if you can use this study strategy. When you start up again, I'll show my answer." We did not capture any data on whether students actually paused the video or tried enacting the study strategy.

**Table 1.** Operationalization of the 6 steps in the Pressley and Harris [9] model.

| Step |
|---|
| **1. Introducing the strategy**<br>The first strategy we'll show you for studying biology is looking for similarities and differences in diagrams. Your textbook shows a lot of material in diagrams, and not all of that information is in the text. Most diagrams have multiple parts, such as different forms of a molecule or different views of the same molecule, chemical reactions, different parts of an organism, or different molecules that are similar in some ways and different in other ways |
| **2. Explaining the usefulness of the strategy**<br>Paying close attention to the similarities and differences across complex diagrams can help you better understand the biology and will definitely help you on exams. Studying the diagrams doesn't just help some types of students, it can help all students |
| **3. Demonstration of how to enact the strategy**<br>I'm going to show you how I would do comparing and contrasting in diagrams in Chap. 27. I'm looking at Fig. 27.3 on p. 569. *Gram staining a) gram-positive bacteria. B) gram-negative bacteria.* The photo doesn't help much, but for gram-positive we've got *cell wall, plasma membrane*, and the cell wall is a *peptidoglycan layer* and the plasma membrane is a.… plasma membrane. The gram-negative also have the cell wall and plasma membrane, but the cell wall is different, it has an *outer membrane*, then a thin peptidoglycan layer at the base of the cell wall. What does staining have to do with it? *Too large to pass through the thick cell wall…masks the safranin dye*. Then on the right, *can pass through this thin cell wall….stains the cell pink or red* |
| **4. Opportunity for practice**<br>Now you try it. Can you use comparing and contrasting in the Fig. 27.11 on p. 573? Pause the video and see if you can use this study strategy. When you start up again, I'll show my answer |
| **5. Feedback**<br>Here's how I did it, yours might be slightly different. *A phage infects a bacterial cell*" so I can see the phage on the outside of the bacterium… |
| **6. Attribution to strategy use**<br>Doing this comparing and contrasting helped me see that every little detail in every diagram is important, and so is all of the text. The color change is pretty small, so I didn't see it until I read through the whole diagram, but now that I know what changes from step 1 to step 5 I can see how the color reinforces that |

Note: *Italics* indicate parts of the script that are read directly from the diagram.

Note that the compare-and-contrast strategy is a multi-faceted one, combining feature detection (‘the phage’), spatial relations (‘on the outside’), comparison processes (‘The gram-negative also have the cell wall and plasma membrane’), contrasts (‘but the cell wall is different’), deductive inferences (‘thick cell wall…too large to pass through…masks the dye…thin cell wall…can pass through and stains), and uses of conventional diagram features (‘I can see how the color reinforces that’).

After the scripts were checked for biological accuracy and pedagogical soundness by the second author, we videotaped the scripts using Camtasia screen capture to record the diagram, pointing the mouse at relevant parts while reading the script. This was then captioned and produced as an.mp4 video and posted to the study Blackboard site via a link to the Ensemble video platform to allow viewing via web link, preventing download and sharing of the videos across conditions.

## 2.4 Procedure

Participants signed a paper consent form and scheduled a proctored session in a computer lab to complete online demographics and various pre-intervention cognitive and motivational measures not described here. After that 1-h session, all further engagement was done via the study-specific Blackboard site at a time and place convenient to the student. After completing the pretests, students were randomly assigned by Blackboard to conditions including study strategies, conditions including worked examples (not described here), and a no-treatment control condition. Assignment to a condition restricted each student’s access within Blackboard to only those intervention components they had been assigned to. Weekly reminder emails for specific conditions were sent to remind participants to access the intervention materials.

Once pretests were completed around the 2nd week of the semester, we launched the Compare-and-contrast in diagrams videos for the first two chapters (already taught in class) and then on the Sunday before each subsequent chapter was taught in lecture. Video access was at a lower rate in the weeks before the exam and reached a peak on the day before the exam.

Despite emailed reminders, before Exam 1 only 34 students had accessed the strategy videos, out of 59 assigned to study strategies. We therefore analyze the effect of these videos on exam scores of the students who actually did access them, compared to all students who did not access them (either the 59 assigned to worked examples, 12 control students, or the 25 non-compliant students [3 students of 131 assigned to conditions did not take Exam 1]).

## 2.5 Data Analysis

After verifying that exam scores were normally distributed within groups and that there were equal variances across groups, we conducted a one-tailed (CC DIA > no CC DIA) independent-samples *t* test to compare exam scores of those who accessed Compare-and-contrast in diagrams videos to those who did not. All analyses were conducted using SPSS 24.

# 3 Results

## 3.1 Differences Between Those Accessing CC DIA and not Accessing

Students accessing the CC DIA videos before exam 1 scored statistically significantly higher on the exam ($M = 79.06$, $SD = 15.03$) than those not accessing ($M = 73.57$, $SD = 16.02$; $t$ [126] = 1.74, $p = .042$, $d = .35$). This advantage is not only statistically significant, but represents an increase from an average B to a B+ using the course grade rubric.

# 4 Discussion

## 4.1 From Basic to Applied Research on Learning from Diagrams

Our brief (approx. 40 min over 4 weeks) video-based instruction on the Compare-and-contrast in diagrams strategy was associated with significantly better scores for the exam on the material covered in the videos. It appears that modeling the complex process of comparing-and-contrasting within diagrams can substantially help students learn biology content. This study strategy is multifaceted, as it incorporates noticing parts of the diagram (selective attention), using conventions such as color coding and arrows, comparing within diagrams, contrasting within diagrams, deductive reasoning (e.g., "What does staining have to do with it?" in Table 1), summarizing at the end of each diagram, and modeling metacognitive monitoring strategies. In this way, it is quite similar to other multifaceted strategy instruction programs that have been tried with multiple representations [11]. Despite this complexity, viewing brief examples that model the process of comparing-and-contrasting in biology is effective in increasing student learning. This is particularly impressive, as the exam did not test any of the instructional materials directly, so we can consider it a measure of near transfer.

Like the small number of other instructional programs implemented with diagrams and other visualizations [12], our intervention was successful at improving learning. Such applications of laboratory findings—that experts engage in comparing and contrasting within diagrams—to classroom experiments continue to serve an important role in providing improved instruction.

# References

1. Guthrie, J.T., McRae, A., Klauda, S.L.: Contributions of concept-oriented reading instruction to knowledge about interventions for motivations in reading. Educ. Psychol. **42**(4), 237–250 (2007)
2. Kok, E.M., de Bruin, A.B., Robben, S.G., van Merrienboer, J.J.: Learning radiological appearances of diseases: does comparison help? Learn. Instr. **23**, 90–97 (2013)
3. Maier, J., Richter, T.: Fostering multiple text comprehension: how metacognitive strategies and motivation moderate the text-belief consistency effect. Metacogn. Learn. **9**(1), 51–74 (2014)

4. Wineburg, S.S.: Historical problem solving: a study of the cognitive processes used in the evaluation of documentary and pictorial evidence. J. Educ. Psychol. **83**(1), 73–87 (1991)
5. Rouet, J.F., Favart, M., Britt, M.A., Perfetti, C.A.: Studying and using multiple documents in history: effects of discipline expertise. Cogn. Instr. **15**(1), 85–106 (1997)
6. Schwartz, D.L., Bransford, J.D.: A time for telling. Cogn. Instr. **16**(4), 475–522 (1998)
7. Alfieri, L., Nokes-Malach, T.J., Schunn, C.D.: Learning through case comparisons: a meta-analytic review. Educ. Psychol. **48**(2), 87–113 (2013)
8. Cromley, J.G., Snyder-Hogan, L.E., Luciw-Dubas, U.A.: Cognitive activities in complex science text and diagrams. Contemp. Educ. Psychol. **35**, 59–74 (2010)
9. National Research Council. A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. National Academies Press, Washington, DC (2012)
10. Pressley, M., Harris, K.R.: Cognitive strategies instruction: from basic research to classroom instruction. In: Handbook of Educational Psychology. Erlbaum, Mahwah (2006)
11. Schlag, S., Ploetzner, R.: Supporting learning from illustrated texts: conceptualizing and evaluating a learning strategy. Instr. Sci. **39**(6), 921–937 (2011)
12. Renkl, A., Scheiter, K.: Studying visual displays: how to instructionally support learning. Educ. Psychol. Rev. **29**(3), 599–621 (2017)

# The Role of Top-Down Knowledge in Spatial Cueing Using Hierarchical Diagrams

Takeshi Sugio(✉)

Doshisha University, 1-3 Tatara-Miyakodani, Kyotanabe, Kyoto, Japan
`tsugio@mail.doshisha.ac.jp`

**Abstract.** The way in which cueing a local element of a hierarchical diagram influences the distribution of visual attention was examined. Using a modified spatial cueing paradigm, the relation between the cue and the target was manipulated by two factors: the level at which the target was presented (higher, identical, lower), and the component to which the target belonged (same or different). The results showed an interaction between these two factors, and simple main effect analysis revealed that the detection time of the target was influenced by three factors: belongingness to the same component, geometrical collinearity of the nodes, and top bias, which regards the top of the diagram as being more informative. All of these factors are related to the conventional knowledge normally possessed about the particular category of the diagrams, and to how such knowledge affects the efficiency of the diagram comprehension process.

**Keywords:** Hierarchical diagram · Visual attention
Conventional knowledge

## 1 Introduction

### 1.1 The Global Structure and Local Components of Hierarchical Diagrams

Schematic diagrams are essential tools for visual thinking, and hierarchical diagrams are representative examples of such diagrams [1]. They represent graphically the hierarchical relationship between items as a difference in levels, which are visually salient. In addition, they consist of the nodes that indicate the objects in the real world, and the line segments that connect each node to another at an either higher or lower level in the hierarchy. Some examples of hierarchical diagrams include: phylogenetic trees that show the evolutionary relationship between different plants and animals, tournament brackets that are often seen in sporting events, such as the Olympics, and organization charts that illustrate the relationships between the parts of and positions within an organization, such as a company. As seen in such examples, the levels in hierarchical diagrams may

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 500–508, 2018.
https://doi.org/10.1007/978-3-319-91376-6_45

represent abstract relations in both the spatial and temporal dimensions. Due to the visuospatial nature of such diagrams, we mistakenly feel that we can understand them intuitively without any prior knowledge in comparison with verbal description of the same data.

It has often been reported that visual attention is first directed to the global properties of stimuli, a phenomenon referred to as global precedence [2]. Therefore, when first glancing at a diagram, it is likely that we extract its global properties. The globality of the diagram's stimulus properties can be defined in terms of its level in the hierarchy of the data. In general, diagrams are more abstract than pictures [3]. Although they need not resemble the objects that they depict, the spatial and temporal relationships among the objects should be mirrored in the visuospatial structure of diagrams. Such relational information can be derived from multiple local elements of diagrams, so it occupies a higher level in the information hierarchy. Owing to the nature of diagrammatic representation, we can use relational information within the diagram to predict the possible behavior outcomes of the objects.

What kind of relational information in the diagrams, and how it is represented, differs according to the categories to which they belong [1]. For example, a matrix-type table represents combinatorial information between two different sets of items in an exhaustive manner, which is to say that the global structure of matrices is invariant with respect to changes in the actual existence of respective relations. If the problem that needs to be solved demands a close examination of the possible combinations of two sets of items, depicting the situation as a matrix is useful. Meanwhile, to obtain some information from the matrix, we need to have conventional knowledge of how it is constructed. By using such knowledge, we can guide our attention to the task-relevant relation effectively.

For hierarchical diagrams, their global shapes can be characterized in terms of the top and bottom sides of a circumscribed rectangle. Since the number of items decreases as the level in the hierarchy becomes higher, it is normally considered that a shorter side depicts a higher level. This property is characteristic of hierarchical diagrams in comparison with other types of schematic diagrams (i.e., matrices and networks, [1]). Whether such a property may act as a retrieval cue for the category to which the diagram belongs depends on the conventional knowledge of the observer.

On the other hand, the local components of the hierarchical diagrams need to be defined with care. The main purpose of reading a hierarchical diagram is not only to know what items are represented in the diagram, but also to obtain information about the hierarchical relationship among them. Therefore, neither a single node nor a line segment connecting multiple nodes may function as a component of a hierarchical diagram. According to Novick and Hurley, a building block of the hierarchical diagram consists of at least three nodes and two directional links connecting these nodes [1]. Such a block indicates the minimal unit of relational information. Thus, to obtain information from a hierarchical diagram in an effective way, the application of conventional knowledge is inevitable. Although appropriate knowledge can be activated in various ways, the global structure of the diagram serves as an effective cue for retrieval.

### 1.2 Effects of the Conventional Knowledge on Spatial Attention to Hierarchical Diagrams

As we have seen, conventional knowledge of a particular type of diagram plays an essential role in both its construction and comprehension. Empirical evidence showing the interaction between reading tasks and graph formats is compatible with this view [4]. The interaction between the bottom-up and top-down processes facilitates robust and efficient diagram comprehension. Trapp and Bar reviewed empirical findings showing the competitive nature of the perceptual process, and proposed a hypothesis that expectations were derived from the low spatial frequency component of visual images [5]. In the case of a diagram, conventional knowledge about how the information is visually organized may guide the attention of the viewer in an appropriate way. At the same time, perceptual cues such as the diagram's global shape or salient features may help activate the category to which the diagram belongs. In other words, both the bottom-up perceptual process and the use of top-down, conventional knowledge are involved in diagram comprehension, and understanding the underlying mechanisms of this comprehension may account for the effectiveness of diagrams in many situations.

Thus, when reading a diagram, visual attention is guided by both salient features and top-down knowledge. Understanding the attentional process in reading a diagram is important in that it shows both temporal and spatial range of information processing capabilities involved. There is also an empirical evidence that spatial attention is required for semantic processing [6]. Furthermore, studying the attentional mechanism of diagram reading could enable us to obtain a better understanding of how the diagram should be designed. However, few experimental results about such a mechanism are available to date. In the present experiment, a modified version of the spatial cueing task was used to examine the conditions in which the benefits for cueing arise [7].

## 2 Experiment

The purpose of the present experiment was to examine how cueing a local element of a hierarchical diagram affects the distribution of visual attention in the diagram spatially. Abstract four-layer hierarchical diagrams were used as stimuli (Fig. 1). Participants were told to fixate on the center of the diagram, and their task was to detect changes in the luminance of the rectangle, which represented a particular item. Before the change in luminance, one of the rectangles was brightened to induce the orientation of visual attention. In valid trials, the change in luminance occurred at the cued rectangle. In invalid trials, the change in luminance took place at a non-cued rectangle. The differences between the cue and target, that is, the rectangle at which the luminance was changed, were manipulated with respect to the following two variables: target level in regard to the cue (higher, identical, lower), and belongingness to the same component (a V-shaped figure that consists of three nodes connected by directional lines). Since the distance between the cue and the target is averaged across conditions, any performance difference between the conditions should be attributed to the differences in the objects that participants form internally with the stimuli. Within the same object, attention directed at some part of it automatically spreads, resulting in benefits in performance at positions other than the cued position.

## 2.1 Method

*Participants.* The participants were 20 undergraduate students at Doshisha University (five males and 15 females, mean age 21.9 years). All students were all paid for their participation, and all had normal or corrected vision.

*Apparatus.* Stimulus presentation and data collection were managed by SuperLab version 5 (Cedrus Corporation) running on a personal computer (HP Compaq Elite 8300SFF) with 17-inch cathode ray tube monitor (Iiyama, HF703U). Screen resolution was set to 1280 × 1024 pixels during the experiment. The viewing distance was set as about 60 cm, and the participants' heads were stabilized using a chin rest. Responses were measured using a RB-530 response pad (Cedrus Corporation).

*Stimuli.* The stimuli were abstract hierarchical diagrams with rectangles as nodes (Fig. 1). All diagrams were composed of four layers, and each node was connected to two other nodes at a lower level. The fixation cross, stimuli, and target were gray, and the cue was white. The fixation cross was a plus sign, which subtended about 0.04° × 0.04°. Each stimulus subtended about 10° × 10° with a stroke of approximately 0.02°, and a square node subtended about 0.04° × 0.04°. The retinal distance between the cue and the target in the invalid condition was about 2.5°. The retinal distance between the fixation cross and the target was 2.5° on average, ranging from approximately 1.0° to 5.0°.

*Design.* The target appeared at either the same square as the cue (valid condition) or a different square from the cue (invalid condition). In the invalid condition, the target appeared at one of six different locations, that is the combinations of the level in the hierarchy compared to the cue (high, identical or low) and the belongingness to the same component as the cue (same or different). In total, 60% of the trials were valid, and 24% were invalid (4% for each condition). The remaining trials were catch trials in which no targets appeared.

*Procedure.* Each trial began with the presentation of overlapped stimuli (a fixation cross) for 1,000 ms, after which, the cue, brightening (i.e., a change in the luminance of the square from gray to white), was superimposed for 100 ms, and then returned to gray. The fixation display was presented again for 200 ms, and the target was presented (the square was filled-in). The target remained on the screen until a response was given, or, if there was no response, for 2,000 ms. Participants were told to respond by pressing the center button of a response pad as quickly as possible when they detected the target, and to withhold responses in the catch trials. A subsequent trial began after a 500-ms interval. The hierarchy

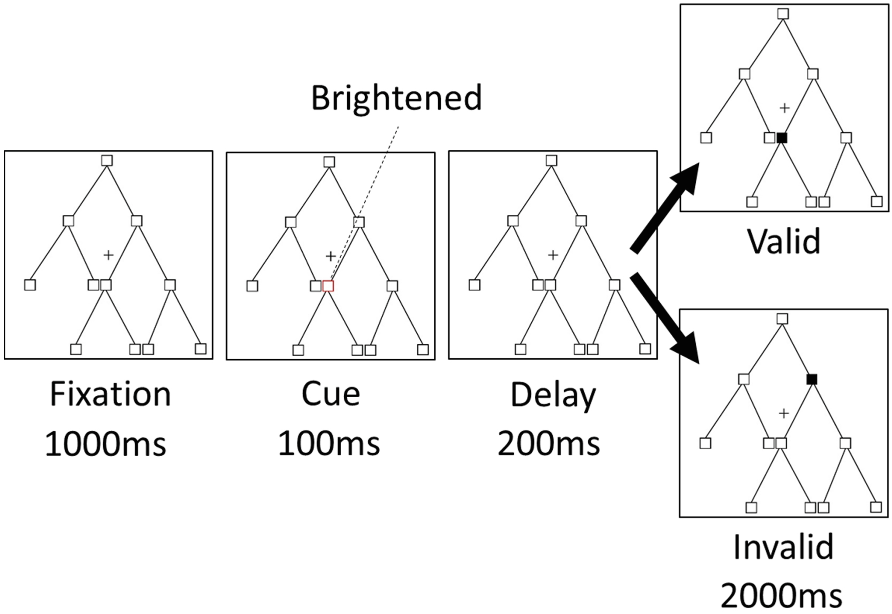


**Fig. 1.** The time course of a single trial in the experiment.

level at which the cue appeared was counterbalanced across trials to control the difference in density between above and below the cued level. The presentation sequence was randomized across participants. The time course of a single trial is shown in Fig. 1.

The participants were told that although their response latency would be recorded, it was important to minimize the number of errors. If a participant made an anticipatory response, defined as a response within 150 ms of the target presentation or a false alarm, a feedback beep was presented for 500 ms. Participants were also asked to maintain their focus on the fixation cross throughout each trial.

The experiment consisted of eight blocks of 125 trials each. Participants were allowed to take a rest between blocks if necessary. The task was explained at the beginning of each block and before the main trials, and 27 practice trials were given. If a participant could not respond correctly for 20 consecutive trials, a practice session was restarted, and the task was explained again. When there was no response to the target within 2,000 ms, when a response was made within 150 ms of the target response, or when the participant failed to withhold a response for the catch trial, the trial was considered as an error.

## 2.2 Results

The total error rate was 1.0%. The participants' median response latencies for correct trials under both the valid and invalid conditions were then compared using a paired t test; the results showed a significant difference (valid condition = 411.350 ms vs. invalid condition = 461.325 ms; $t(19) = 2.330, p = .031$).

The median response latencies for the invalid conditions were analyzed using repeated analysis of variance (ANOVA), with the within-participant factors of target level and component. The results are shown in Fig. 2. The main effect of target level and the interaction between two factors were significant (Greenhouse-Geisser adjusted results for target level: $F(1, 19) = 4.055, p = .045, \eta_p^2 = .176$; interaction: $F(2, 38) = 4.318, p = .020, \eta_p^2 = .185$). Multiple comparison for target level performed using Shaffer's method found that the lower condition took significantly longer than the higher or identical conditions (both $p < .05$). As for the interaction, simple effect analysis revealed that the effect of level was marginally significant when both the cue and the target belonged to the same component (Greenhouse-Geisser adjusted $F(2, 38) = 3.454, p = .058, \eta_p^2 = .154$), and highly significant when the target appeared at the node of a component different from the cue ($F(2, 38) = 5.827, p = .006, \eta_p^2 = .235$). Multiple comparison for the simple effect under the same component condition showed a significant difference between the lower and identical levels ($p < .05$), and that the higher-level condition was significantly faster than the lower- and identical-level conditions for the different component conditions (both $p < .05$). The simple effect of the component was only significant when both the cue and the target were presented at the identical level in the hierarchical diagram ($p < .05$).

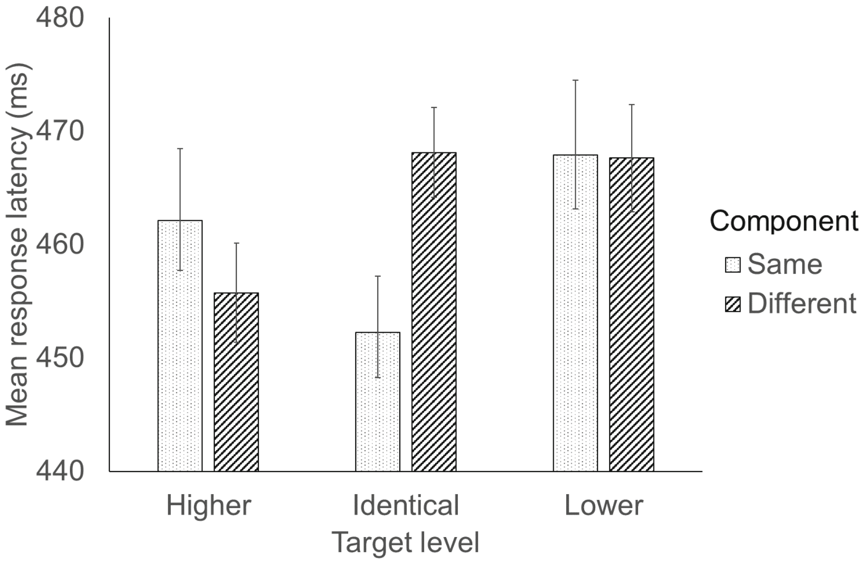


**Fig. 2.** Interaction between the target level and component for response latencies.

## 2.3 Discussion

The results clearly demonstrate that the effects of spatial cueing of a node in the hierarchical diagram differ according to the hierarchy level at which the target appears and the local V-shaped component to which the target belongs; both factors determine the informational structure of the hierarchical diagram. The present results support the claim that when viewing a hierarchical diagram, conventional knowledge about how the diagram is organized globally, and what constitutes its elements is activated automatically. Consequently, task-relevant information can be obtained from the diagram efficiently.

The main effect of target level showed that when a target appeared at a level lower than the cue, more time was needed to detect the change in luminance. This was not the case for the higher level, which suggests that we read the diagram from a level higher than the cued level in the hierarchy preferentially. Based on the global shape of the hierarchical diagram, we can easily determine which side is the top. We know that the top side of a hierarchical diagram depicts the items that indicate superordinate concepts, and such biases might help us comprehend the diagram.

The interaction between the target level and component is also important in that it suggests how nodes are organized as building blocks across different levels in the hierarchical diagram. When the target level was different from the cue, no significant difference was observed in detection time between the component conditions. On the other hand, when the target level was identical to the cue, the target was detected faster under the same component condition. This suggests that attending to a particular node in a diagram benefits from both the global structure and the local component of that diagram. Identifying the level in the hierarchy is essential in determining the representational range of the node, e.g., whether the node represents an animal or a fish. It may also be related to the entry level often discussed in the categorization literature [8]. By using the global features of a diagram, such as node collinearity, we may be able to identify the hierarchical structure and the level considered the entry point.

The examples in which the target was detected relatively faster are shown in Fig. 3. When the target was on the identical level as the cue and belonged to the same component (the left panel in Fig. 3), detection was faster, suggesting that spatial cueing is affected in two distinct manners. On the one hand, visual attention is directed at the information unit of the diagram, which consists of three nodes connected by two directional line segments. When one of the nodes is cued, the effect automatically spreads to other nodes in the same unit. On the other hand, visual attention is directed at a particular level in the hierarchy; this is guided by geometrical features such as collinearity. However, as shown in the right panel in Fig. 3, when the target appears at the node of a different unit, the node for which the level is higher than the cue is detected faster. These results suggest that the effect of geometrical collinearity is contingent on the information unit to which the node is assigned. If the target node does not belong to the same information unit as the cue node, the higher level is detected faster, indicating that the reading process takes place from a particular direction in the hierarchy

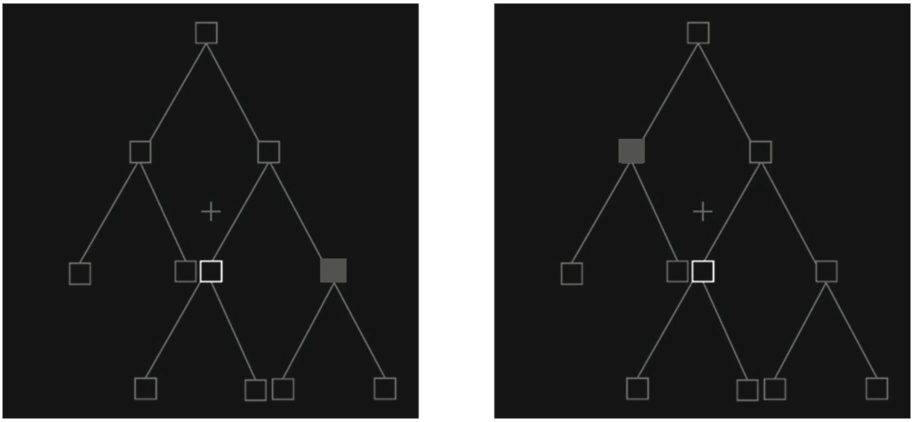


**Fig. 3.** Examples of the stimulus for which the detection of the targets (filled squares) was relatively rapid (the cues are the squares with brighter contours).

("top bias", [9]). In order to resolve this confounding, the follow-up experiment was performed by using the same stimuli in the inverted position. According to my unpublished data from this follow-up experiment, the simple effect of level for the interaction was significant only for the same component condition. This result suggested that when the top of hierarchy in the diagram is consistent with that of a visual scence, the viewer receives attentional benefits both from the component of the diagram and the environmental upright. This attentional benefit from the environmental upright eliminates when the hierarchical diagram is presented in the inverted position, but the effect of the component remains.

In sum, the following three different types of spatial features are used in reading a hierarchical diagram to implement an efficient process: belongingness to the component, geometrical collinearity, and top bias. The usefulness of these features is based on the conventional knowledge of hierarchical diagrams that has been acquired throughout encounters with them in daily situations.

## 3 Conclusion

The present study examined how the global structure and local components of a hierarchical diagram influence the cueing effect on a particular node in the diagram. How a particular diagram is constructed and comprehended depends on the conventional knowledge possessed by the observer; the results suggest that different types of perceptual features influence the orientation of visual attention to the encountered diagram. These results might also provide a clue to the appropriate design for diagrams used in a variety of situations in daily life.

The use of conventional knowledge requires a certain amount of processing resources, such as working memory. Hence, individual differences in working memory might be related to performance in a comprehension task. Furthermore, it has been reported that individuals with autism spectrum disorder require more time for global processing [10]. How such individual differences affect performance in a diagram task might contribute to the concept of universal design, and should therefore be examined in a future study.

**Acknowledgments.** I would like to thank Naoko Takada for conducting the experiment. This work was supported by JSPS KAKENHI Grant Numbers 15K00215 and 15H02716.

## References

1. Novick, L.R., Hurley, S.M.: To matrix, network, or hierarchy: that is the question. Cogn. Psychol. **42**(2), 158–216 (2001)
2. Navon, D.: Forest before trees: the precedence of global features in visual perception. Cogn. Psychol. **9**(3), 353–383 (1977)
3. Reed, S.K.: Thinking Visually. Psychology Press, New York (2010)
4. Shah, P., Freedman, E.G., Vekiri, I.: The comprehension of quantitative information in graphical displays. In: Shah, P., Miyake, A. (eds.) The Cambridge Handbook of Visuospatial Thinking, New York, NY, pp. 426–476. Cambridge University Press (2005)
5. Trapp, S., Bar, M.: Prediction, context, and competition in visual recognition. Ann. N. Y. Acad. Sci. **1339**(1), 190–198 (2015)
6. Waechter, S., Besner, D., Stolz, J.A.: Basic processes in reading: spatial attention as a necessary preliminary to orthographic and semantic processing. Vis. Cogn. **19**(2), 171–202 (2011)
7. Egly, R., Driver, J., Rafal, R.D.: Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. J. Exp. Psychol. Gen. **123**(2), 161–177 (1994)
8. Rosch, E., Mervis, C.B., Gray, W., Johnson, D.M., Boyes-Braem, P.: Basic objects in natural categories. Cogn. Psychol. **8**(3), 382–439 (1976)
9. Chambers, K.W., McBeath, M.K., Schiano, D.J., Metz, E.G.: Tops are more salient than bottoms. Percept. Psychophys. **61**(4), 625–635 (1999)
10. Van der Hallen, R., Evers, K., Brewaeys, K., Van den Noortgate, W., Wagemans, J.: Global processing takes time: a meta-analysis on local-global visual processing in ASD. Psychol. Bull. **141**(3), 549–573 (2015)

# Exploring the Relationship Between Visual Context and Affect in Diagram Interpretation

Matthew Wood[1,2](✉) and Susan Stocklmayer[1]

[1] Australian National Centre for the Public Awareness of Science,
Australian National University, Canberra, Australia
`mattwood@biol.tsukuba.ac.jp`,
`sue.stocklmayer@anu.edu.au`
[2] Faculty of Life and Environmental Sciences,
University of Tsukuba, Tsukuba, Japan

**Abstract.** Among the growing body of research on the interpretation of diagrams there appears to have been relatively little attention paid to emotional or attitudinal responses, despite the fact that they may be significant for communicators aiming to stimulate interest, influence attitudes, or motivate action. This research explores the impact on affect of visual context in biological life cycle diagrams. In two qualitative studies, participants viewed decontextualized life cycle diagrams along with diagrams that included a contextual backdrop, and discussed their interpretations, associations, and attitudes toward the diagram content. Thematic analysis of the data revealed that context was associated with an elevated sense of empathy and concern for the animal, and a stronger perception of personal relevance—clear indications that diagram design can have important emotional and attitudinal impacts.

**Keywords:** Context · Diagram · Affect · Empathy · Relevance

## 1 Background and Introduction

Science graphics are studied as tools for conveying data, conceptualizing science, and teaching science, however they are less studied as tools for communicating science and related issues to non-specialist audiences—a common use of diagrams in science communication and environmental communication settings.

The majority of existing research on the design of science diagrams (and perhaps diagrams in general) focuses on cognitive outcomes such as information transfer, comprehension, recall and so on. These are undoubtedly important outcomes that are of great interest to educators and designers of data and information displays. However this narrow focus on cognitive outcomes ignores a swath of other affective, emotional, and attitudinal outcomes that are potentially important for communicators. For example, in addition to information transfer, environmental and risk communicators may be interested in the emotional impacts of graphics aimed at promoting attitude and behavior change. Similarly, science communicators (and science educators for that matter) may be interested in the influence of a graphic on interest and motivation in addition to its impact on understanding.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 509–516, 2018.
https://doi.org/10.1007/978-3-319-91376-6_46

A common recommendation for instructional graphic design is that clarity and comprehension should be achieved through focus and simplicity, driving a tendency toward reductionism in diagram design. This notion is supported by a variety of research which suggests that recognition, transfer, and understanding are improved by reducing extraneous details [e.g. 1–3]. However, removing the 'extraneous' also often means removing context, and context has been shown to be not just important for learning and motivation [e.g. 4], but fundamentally critical for cognition [5] and meaning making [6].

This paper poses the following question. While there is a significant amount of research to support reductionism in diagram design as an effective approach for information transfer and data interpretation outcomes, what affective roles might visual context play in the interpretation of diagrams that have hitherto been overlooked? With little indication from the literature for an obvious point of departure for such a study, we took a broad exploratory qualitative approach aimed at identifying interesting and potentially important aspects of the relationship between context and affect. Rather than numerical confirmations of the prevalence of a pre-determined relationship, qualitative methods allowed insight into the spectrum of possibilities that we could not have known otherwise. This was particularly suitable for the exploratory nature of this project. Specifically, this project investigated impacts of visual depictions of environmental contexts in biological life cycle diagrams on viewers' attitudes toward the subject animal.

## 2 Study 1

To explore potential influences of visual context on viewers' interpretation and attitudes toward content, a series of focus groups was conducted at the University of Tsukuba, Japan. The study did not target any particular aspect of this relationship, but rather aimed to capture the widest possible range of responses to context.

### 2.1 Method

Three 90-min focus groups were conducted involving a total of 17 participants (19–39 yo) of various nationalities and backgrounds. This cultural and demographic heterogeneity was expected to encourage a rich diversity of responses. Each focus group consisted of a word-association exercise and a loosely guided discussion about participants' interpretations, associations, and attitudes regarding a selection of life cycle diagrams with and without a contextual background.

A total of five diagrams portraying fish (salmon, snapper) and crustacean (prawn, lobster) life cycles were sourced from websites and online publications for use in the focus groups. Diagrams were categorized as either contextualized or decontextualized. Decontextualized diagrams were completely devoid of any background or visual context, and simply represented the various life stages of the animal in a circular arrangement on a plain white backdrop. Contextualized diagrams showed a similar circular arrangement of life stages, but on a backdrop depicting the habitat of the animal, including different habitats for different life stages. The visual context also included direct or indirect indications of the presence of humans (e.g. nearby houses, a boat, a scuba diver). The subject species were selected to be less likely to invoke strong emotive responses (e.g. cute, repulsive).

Focus group sessions were recorded and later transcribed for analysis using the software package, NVivo 11 by QSR International. Thematic analysis of the transcriptions was conducted using a Template Analysis approach [7].

## 2.2 Themes and Outcomes

This section describes the major themes to emerge from the focus group data.

*Context and Habitat*

The contextualized diagrams depicted species within their natural habitats, and this prompted many participants to view those animals as free, living entities embedded within their natural environments. This was in stark contrast to impressions that decontextualized diagrams represented dead animals. Participants were clearly linking the animal with where it lives at different stages in its life, and this in turned raised further questions for some participants about how and why the animal moved, its dependence on its environment, and the inter-relatedness of the various habitats.

*Context and Empathy/Concern*

Of the 15 participants who were asked directly if they felt differently about the species represented in a contextualized life cycle diagram compared to the species represented in a decontextualized diagram, 12 responded that they did. These participants indicated that they viewed the contextualized species more as a living organism, and were more able to understand how the animal lives in its natural environment. During the discussion, participants clearly indicated more emotional investment in species depicted with a contextual background. This often emerged as a tendency to view the animal's life from the point of view of the animal itself, greater sensitivity to the challenges the animal faces throughout its life, as well as direct expressions of empathy.

*Context and Vulnerability*

Some participants commented that they saw species represented in contextualized diagrams as vulnerable to environmental change or to human influence. In some cases this was prompted by indications of the presence of humans in the diagram.

## 2.3 Discussion

Diagram context certainly appeared to have some effect on viewers' attitudes and opinions toward the content. For many participants visual context encouraged a view of the animal as an integrated part of its environment, and invoked feelings of empathy and concern for the species portrayed.

Environmental educators have stressed the need to promote empathy for the natural world if conservation efforts are to be successful [e.g. 8]. Environmental psychology researchers have found empathy to be a reliable predictor of conservation behavior [9], and have shown that under some conditions, environmental concerns could be influenced by manipulating an individual's sense of empathy for the animals involved [10]. Perhaps evident of this relationship between empathy, environmental concern, and environmental behavior, in this study context was not only associated with empathy, but also seemed to prompt participants to consider stresses on the animal from environmental and fishing issues, to contemplate the importance of habitat conservation, and to imagine how the animal lives its life from the animal's own perspective. It follows that a relationship between visual context and empathy as observed in this study would be of interest to environmental psychology research and potentially of value to environmental communicators.

*Limitations*

Although the diagrams used in this study were carefully selected to represent the contextualized and decontextualized archetypes, there was considerable variation in color, illustration style, and aesthetics. This variety contributed to a richer range of responses to the diagrams, as was intended, but also impeded a strict comparison of context versus no context. Consequently, it should be acknowledged that when participants responded in a particular way to context, they may also have been responding to some extent to color, illustration style, or other aesthetic qualities that could not be considered separately from context in this study.

## 3 Study 2

To further confirm and explore the findings of Study 1 and to address a number of limitations in that study, a second series of focus groups was conducted. In this second study, diagrams were purpose-built for more control over design and a more direct comparison of the impact of context. Also, the questioning route was more focused towards recognition of connections between the animal and its environment, as well as feelings of empathy and concern for the animal.

### 3.1 Method

Four 90-min focus groups were conducted at the Australian National University, Australia, with a total of 14 participants (18–33 yo). Again, participants represented a variety of nationalities and backgrounds, increasing the potential diversity of responses. The association exercise was conducted using the decontextualized life cycle diagram of one of the species used in this study, and then repeated with the contextualized diagram of the other species (i.e. each focus group used two diagrams—either MA-1/BB-1 or BA-1/MB-1).

The diagrams used in this study (Fig. 1) were generalized life cycles of an aquatic mussel and a gooseneck barnacle. These species were chosen for their expected unfamiliarity, relative complexity of their life cycle, and low likelihood of invoking an extreme emotional reaction. Contextualized and decontextualized versions contained identical representations of each animal's life cycle, differing only in the presence or absence of an illustrated background depicting the animal's environment. Each life stage was labeled and short text descriptions were added for further explanation of the life cycle and to ensure both versions contained details on habitat location (also presented visually in the contextualized diagram).

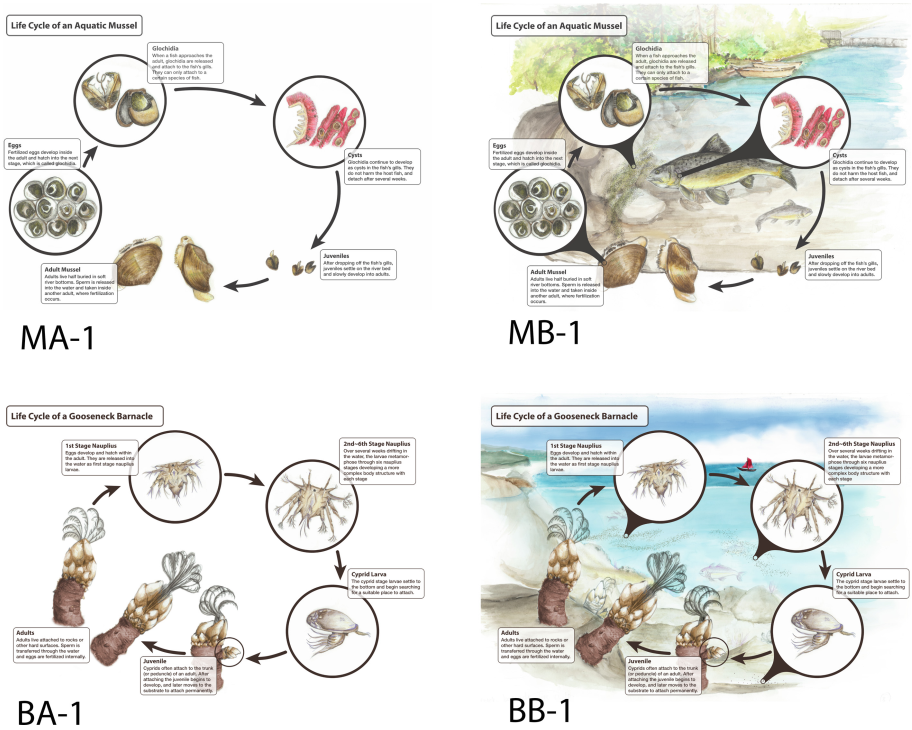


**Fig. 1.** Diagrams used for Study 2 focus group discussions: MA-1 decontextualized life cycle of an aquatic mussel, MB-1 contextualized life cycle of an aquatic mussel, BA-1 decontextualized life cycle of a gooseneck barnacle, BB-1 contextualized life cycle of gooseneck barnacle.

Each focus group session was recorded, and later transcribed and analyzed with NVivo 11 software by QSR International. As in Study 1, data analysis was guided by a template analysis approach.

### 3.2 Themes and Outcomes

The most significant themes to emerge from Study 2 were as follows.

*Context and Relevance*

Although participants once again clearly indicated feelings of empathy for the animals depicted in diagrams, and concern about their fate, a much more pronounced theme to emerge from Study 2 was a strong and clear indication that context led participants to feel a greater sense of personal connection and involvement with the animal depicted and the environment it lives in. There were clear indications of this for both of the species involved in this study. For these participants, recognizing familiar scenes, places, or objects in the contextual background allowed them to feel that the animal, its life, and the information presented in the diagram were more closely connected to their own lives, and that they might encounter that animal or its habitat. We interpreted this as an elevated sense of personal relevance. Conversely, diagrams without context were dismissed as distant, unrelated, and uninteresting.

*Context and Curiosity*

Since the questioning route directly asked participants what else they would like to know about the animal depicted in the diagram, all groups produced comments about aspects of the life cycle that participants were curious about, or that they did not completely understand, and therefore little can be surmised about the mere presence of such comments. There were, however, major differences in expressions of indifference or disinterest (and consequently a lack of curiosity) between diagram types. The questioning route did not directly ask about disinterest, so these types of comments were spontaneous, and almost exclusively referred to decontextualized diagrams.

### 3.3 Discussion

Thus an important outcome to emerge from Study 2 was the apparent link between context and an elevated perception of personal relevance. Participants were sensing a personal connection triggered by the environment depicted in the contextual background, and this was apparent for both species. The educational psychology and persuasion literature reports that relevance is linked to increased attention, motivation, learning [11], interest [12], and cognition [13]. A relationship between context and curiosity, a subset of interest, was not directly apparent in this study, however there was an inverse relationship between expressions of disinterest in the information content of decontextualized diagrams and perceptions of personal relevance of information represented in contextualized diagrams. Although this is not definitive evidence of the entire context-relevance-interest relationship, it is a strong, albeit indirect suggestion that such a relationship may exist. Certainly this study suggests a context-relevance relationship, while the relevance-interest end is predicted from the literature.

There may also be a connection between relevance and empathy. Environmental psychology researchers consider empathy for the natural world to be closely related to, and mediated by a personal connection to nature, both of which are predictors of environmental concern [9]. Definitions of relevance center around the concept of being closely connected, so a personal connection to nature, which mediates empathy and predicts environmental concern, could alternatively be considered as a view that nature is personally relevant. The connection to nature of environmental psychology is an entrenched, stable attitude, while the personal relevance observed in this study was specific and situational, however there is clear potential for a relationship between relevance, empathy, and concern, and this might explain why these themes have emerged strongly over the two studies reported in this paper.

*Limitations*

Any apparent influence of context in this study may also include influence from specific characteristics of the diagram, illustration style, and content used in this study. This should not rule out cautious extrapolation to effects of context more generally, but care should be taken.

## 4 General Discussion and Conclusions

Repeatability is not a realistic expectation for the intricate settings of qualitative research, but rather each research situation is considered unique and assessed on its own merits. The heterogeneity of Studies 1 and 2, then, serve to offer greater breadth for considering potential relationships between context and affect. The studies were, however, consistent, in that context was associated with affective responses in both studies, and the major themes across the two studies—relevance, empathy, and concern—are conceivably related. Taken together the findings of the two studies suggest that context may play a role in encouraging feelings of empathy and concern, and a perception of relevance, but that other mediating factors are also involved. These factors might include the nature and content of the context, along with illustration style, clarity, perspective, viewer background, content familiarity, external prompts, and so on. These studies have made an initial foray into the relationship between context and affect, but subsequent research is required to further clarify the relationships revealed in this paper, and to determine their prevalence among viewers and across content and illustration style.

As discussed in previous sections, empathy and concern are of interest to environmental psychology studies, and relevance has implications for interest, motivation, and learning. So the findings from this study would be of specific interest to researchers in these areas along with communicators and educators. More generally though, these studies clearly reveal that diagram design can have important emotional and attitudinal impacts on viewers. This prompts one to consider what other aspects of design might have emotional and attitudinal impacts, and conversely, what other emotions and attitudes might be influenced by diagram design. The role of affect in interpretation has potential implications for a broad array of researchers and practitioners in fields such as science communication, risk communication, health communication, education, or any other field where diagrams are employed toward objectives that can benefit from an emotional or attitudinal influence. This is certainly an area that deserves further attention.

**Conflict of Interest:** The authors declare that they have no conflict of interest.

## References

1. Cook, M.P.: Visual representations in science education: the influence of prior knowledge and cognitive load theory on instructional design principles. Sci. Educ. **90**, 1073–1091 (2006). https://doi.org/10.1002/sce.20164
2. Smallman, H.S., St John, M.: Naive realism: limits of realism as a display principle. In: Proceedings of Human Factors and Ergonomics Society 49th Annual Meeting (2005)
3. Mayer, R.E., Moreno, R.: Animation as an aid to multimedia learning. Educ. Psychol. Rev. **14**, 87–100 (2002). https://doi.org/10.1023/a:1013184611077
4. Cordova, D.I., Lepper, M.R.: Intrinsic motivation and the process of learning: beneficial effects of contextualization, personalization, and choice. J. Educ. Psychol. **88**, 715–730 (1996). https://doi.org/10.1037/0022-0663.88.4.715

5. Butterworth, G.: Context and cognition in models of cognitive growth. In: Context and cognition: Ways of Learning and Knowing, pp. 1–13. Harvester Wheatsheaf, Hertfordshire (1992)
6. Klassen, S.: A theoretical framework for contextual science teaching. Interchange **37**, 31–62 (2006). https://doi.org/10.1007/s10780-006-8399-8
7. Brooks, J., McCluskey, S., Turley, E., King, N.: The utility of template analysis in qualitative psychology research. Qual. Res. Psychol. **12**, 202–222 (2015). https://doi.org/10.1080/14780887.2014.955224
8. Sobel, D.: Beyond ecophobia: reclaiming the heart in nature education, 2nd edn. Orion Society, Great Barrington, MA (2013)
9. Tam, K.-P.: Dispositional empathy with nature. J. Environ. Psychol. **35**, 92–104 (2013). https://doi.org/10.1016/j.jenvp.2013.05.004
10. Schultz, P.W.: Empathizing with nature: The effects of perspective taking on concern for environmental Issues. J. Soc. Issues. **56**, 391–406 (2000). https://doi.org/10.1111/0022-4537.00174
11. Eccles, J.S.: Expectancies, values and academic behaviours. In: Spence, J.T. (ed.) Achievement and Achievement Motives: Psychological and Sociological Approaches, pp. 75–146. W.H. Freeman, San Francisco (1983)
12. Schraw, G., Lehman, S.: Situational interest: a review of the literature and directions for future research. Educ. Psychol. Rev. **13**, 23–53 (2001). https://doi.org/10.1023/a:1009004801455
13. Petty, R.E., Brinol, P.: The elaboration likelihood model. In: Van Lange, P.A.M., Kruglanski, A.W., Higgins, T. (eds.) Handbook of Theories in Social Psychology, vol. 1, pp. 224–246. SAGE Publications, London (2012)

# Visualising Latent Semantic Spaces for Sense-Making of Natural Language Text

Ana Šemrov[1], Alan F. Blackwell[1](✉), and Advait Sarkar[1,2]

[1] Computer Laboratory, University of Cambridge, Cambridge, UK
ana.semrov@gmail.com, alan.blackwell@cl.cam.ac.uk
[2] Microsoft Research, Cambridge, UK
advait@microsoft.com

**Abstract.** Latent Semantic Analysis is widely used for natural language processing, but is difficult to visualise and interpret. We present an interactive visualisation that enables the interpretation of latent semantic spaces. It combines a multi-dimensional scatterplot diagram with a novel clutter-reduction strategy based on hierarchical clustering. A study with 12 non-expert participants showed that our visualisation was significantly more usable than experimental alternatives, and helped users make better sense of the latent space.

## 1 Introduction

This design case study explores an increasingly common class of diagram, which represents a statistical model used to explore unstructured, qualitative datasets, such as our example dataset: a snapshot of Wikipedia. We focus on Latent Semantic Analysis (LSA), one of a class of methods that represents words as vectors, where dimensions of the vector space capture aspects of word meaning [1,2]. LSA dimensions have been shown to be good predictors of human meaning-based judgements [3], perform well in tasks based on word similarity [4] and are useful in sentiment analysis [5]. Unfortunately, users do not find it easy to interpret the dimensions extracted from LSA.

Our research therefore investigates whether interactive diagrams can be used to provide a more interpretable mapping between a model created using LSA and the domain content of the vocabulary being mapped, and whether a mapping of this kind can provide an effective basis for sensemaking and exploration.

Conventional quantitative graphs are valuable to experts who are interested in understanding and refining the model. It is not unusual for experts in a domain to invent tools that will assist them in their own tasks, and as a result, we find that statistical visualisation approaches are widespread in the data analytics and natural language processing literature. However, such visualisations may be less valuable to those who are not experts.

Our distinctive approach focuses on presenting the semantic relationships between words, treating the problem as one of diagram design. We visualise

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 517–525, 2018.
https://doi.org/10.1007/978-3-319-91376-6_47

semantic structure using geometric regions that summarise clusters of related words. The user can explore any word group cluster by selecting and "expanding" the view to focus on those words. Exploration of clusters can be recursive, allowing navigation of a semantic hierarchy. Interaction with lower levels of the hierarchy allows the user to explore closely related words, while interaction with higher levels provides a thematic overview of the corpus. We use diagrammatic design cues to communicate these different interpretation opportunities to the user.

We demonstrate through a user study that our system improves the ability of non-expert users to discover groups of related words and assign meaning to dimensions, when compared to two more conventional alternative visualisations.

### 1.1 Related Work

Visualisation of multidimensional datasets has been previously explored. ScatterDice uses scatterplots and scatterplot matrices to represent the dataset [6]. An alternative approach uses parallel coordinates and hierarchical clustering [7], lines are coloured according to the proximity of their corresponding data points in a cluster hierarchy. Other approaches to scatterplot matrices, including density contour, sunflower plots, and density estimations, have been compared [8].

A notable prior design, aimed at improving the understanding of latent semantic spaces, is a flattened network visualisation of the space [9]. A separate network can be displayed for each dimension, where the length of edges between words corresponds to the similarity between those words on that dimension.

Strategies for clutter-reduction have been well-explored. Some taxonomies distinguish between clutter reduction strategies affecting the appearance of individual data points, those spatially distorting the space to displace the data points, and animation techniques [10]. Another survey presents visual aggregation strategies including multidimensional scatterplots, parallel coordinates, star plots, and a model of hierarchical aggregation related to our approach [11].

## 2 LSA Model Construction

The Westbury Lab Wikipedia Corpus [12] was used during development as well as the experiment. This snapshot of the English Wikipedia contains articles as plain text without Wiki markup, links and other non-content material.

After removing stopwords and words occurring in fewer than two documents, we constructed word-document co-occurrence matrix $\mathbf{A}$. Rows correspond to words, columns to documents, and each entry $a_{ij}$ corresponds to the appearances of word $i$ in document $j$. We applied inverse document frequency (IDF) weighting; words appearing in fewer documents were prioritised relative to common words. We then used a standard LSA library [13]. The co-occurrence matrix is factorised using singular value decomposition [14]. The $n \times m$ matrix $\mathbf{A}$ can be written as the product $\mathbf{A} = U\Sigma V^T$, where $U$ is an orthogonal $n \times n$ matrix that recasts the original row (word) vectors as vectors of $n$ derived orthogonal factors; likewise $V$ is an orthogonal $m \times m$ matrix describing the original columns. $\Sigma$ is an $n \times m$

diagonal matrix, whose diagonal entries are ‘singular values’ of the matrix and the columns of $U$ and $V$ are respectively the right and left singular vectors.

The top $k$ singular values, and the corresponding rows and columns from $U$ and $V$, give a $k$-rank approximation for $A$. The word vectors in $U$ can thus be expressed with $k$ dimensions, instead of in terms of every document. The choice of $k$ is task and content dependent [15] and is typically tuned empirically [1]. Using the L-method [16] we found $k = 5$ dimensions sufficient for our corpus.

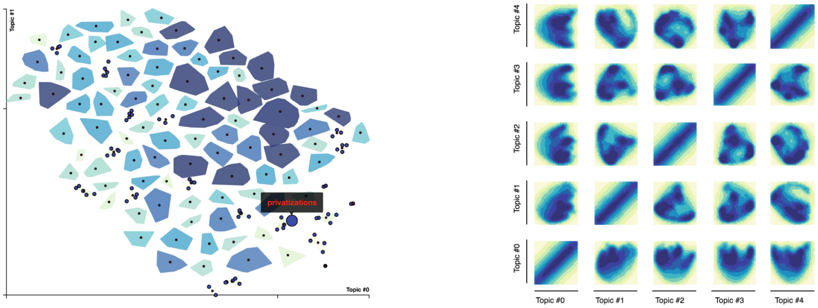
## 3 Interface Design

Our interface consists of: (1) A *hierarchically-clustered* diagram for clutter management, (2) a *heatmap matrix* for navigating between dimensions, (3) a *graphical history* for navigation context, and (4) a *word cloud* for inspecting clusters.

### 3.1 Hierarchically-Clustered Scatterplot

We construct a cluster hierarchy, allowing fluid navigation between multiple levels. This supports *exploratory* analysis, where it is not known in advance which aspects of the data are most important. We use agglomerative hierarchical clustering using Euclidean distance and centroid linkage. Every datapoint is initialised as a separate ‘cluster’. In each iteration, the pair of clusters with the lowest inter-cluster distance is merged. This is repeated until all points have been merged into a single cluster. This process creates a tree (represented as a dendrogram (Fig. 3, left)): the root node is a cluster containing the whole dataset, nodes have exactly two children, and leaves are individual datapoints.

**Visualising a Cluster.** Clustering trades detail about individual data points for aggregate information. A good cluster representation would convey information about its contents (*scenting*) for effective exploration (*foraging*) [17]. In our representation of each cluster (Fig. 1a), the *shape* of the cluster is preserved



(a) Multiple clusters visualised. A word is being inspected via tooltip.

(b) Heatmap matrix. Blue indicates a high probability density.

**Fig. 1.** Cluster visualisation and heatmap matrix. (Color figure online)

by rendering the convex hull of its constituent points; the colour of a cluster is mapped to its *cardinality* – darker clusters contain more points; and, the centroid is plotted in red. Data points (words) are shown explicitly, and can be inspected individually, if a cluster contains very few of them.

**Cluster Expansion.** Double-clicking a cluster expands it, ideally resulting in a display that efficiently uses the available screen space while minimising overlap of the newly displayed clusters. Each expansion may correspond to a descent of multiple levels in the hierarchy tree, based on a criterion that supports the fastest descent of the hierarchy while avoiding clutter.

We developed the heuristic of a 'minimal displayable centroid distance'. The idea is that the centroids of clusters onscreen should never be closer than this amount. We set this to 30 pixels, corresponding roughly to 1 cm on our displays. Clusters higher in the hierarchy tree correspond to a greater centroid distance between that cluster's children. On expanding a cluster, the scatterplot is rescaled to tightly fit the expanded cluster, such that the minimum and maximum value on the $x$ and $y$ axes corresponds to the extents of the cluster on the dimensions being plotted on $x$ and $y$, respectively. The pixel size of the overall scatterplot is known, giving a mapping between data and screen space. From this, we map a distance of 30 pixels back to data space and get the optimal tree cut height (the lowest height where clusters are sufficiently distant) (Fig. 3).

**Fig. 2.** Word clouds corresponding to four clusters. Font size and colour encode the words' distances from the centroid. Can you assign a meaning to each cloud? (Color figure online)

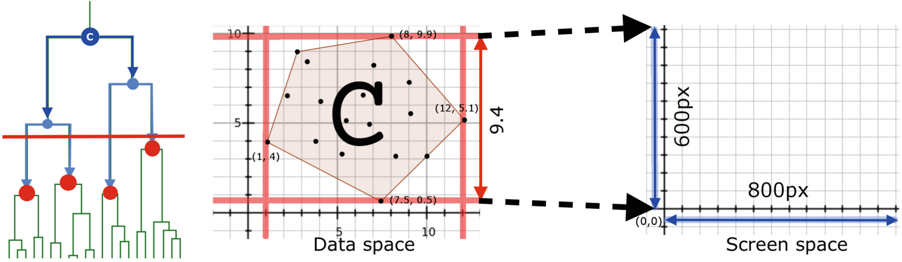


**Fig. 3.** Expanding a cluster. When $C$ is expanded, a cut (red line) is made at the height corresponding to the minimum displayable distance between clusters. (Color figure online)

### 3.2 Heatmap Matrix: Helping Users Navigate Between Dimensions

The user must select which two dimensions of the $n$-dimensional dataset will be plotted. Without guidance, this task can degenerate into tedious enumeration of dimension pairs, or ineffective random switching. A scatterplot matrix displays all dimension pairs, letting users quickly identify plots of interest, but is costly to render: for 30,000 words it requires plotting 30,000 points per dimension pair. One strategy to reduce the rendering cost is to display a naïve random sample, but this only works on uniformly distributed data; with outliers and areas of varying density, it produces distorted or misleading plots.

Our solution is to plot the *sampled probability density* of the data as a heatmap, with colour mapped to density, as seen in Fig. 1b. We used bivariate kernel density estimation (KDE)[18]. This significantly reduces the complexity of rendering while still capturing the overall shape of the data. The resultant heatmap matrix is a navigational aid: users click on cells in this matrix to select which two dimensions are displayed in the cluster diagram.

### 3.3 Word Cloud

There are two ways to inspect words. Hovering on single points displays a tooltip that remains open if the point is clicked. When a cluster is clicked, a subset of the words contained in the cluster is visualised in a word cloud to the right (Fig. 2). To manage the cloud's visual complexity, only the 30 words closest to the cluster centroid are displayed, as these are most representative of the cluster. The size and colour of words are mapped to the distance of the words from the centroid.

### 3.4 Graphical Expansion History

When a cluster is expanded, its place within the larger cluster hierarchy is no longer visible. A graphical history [19] preserves this context. Upon cluster expansion, a snapshot of the current plot, highlighting the expanded cluster,

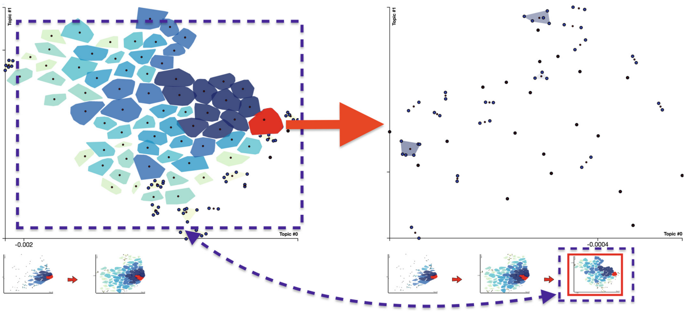


**Fig. 4.** Cluster expansion step. Expanded clusters are marked in red. (Color figure online)

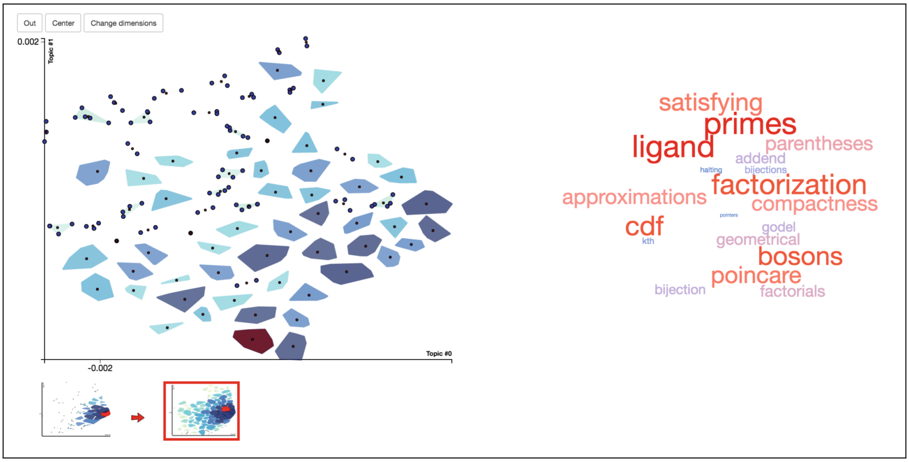


**Fig. 5.** Our interface in use.

is added to the history. A sequence of expansions provides context for each expanded cluster, showing how the expanded cluster relates both to its immediate context as well as the entire data space (Fig. 4). Any snapshot in the history can be clicked to revert to that level, creating a multi-level overview+detail [20].

Taken together, the four components: clustered scatterplot, heatmap matrix, graphical history, and word cloud constitute our interface (Fig. 5). The heatmap matrix is accessed with the 'change dimensions' button, which displays the matrix to the right of the cluster diagram in place of the word cloud.

## 4 User Study

We define two goals of latent semantic space exploration: (1) finding groups of related words and assigning a meaning to the common underlying theme, and (2) interpreting the meaning of each dimension. We were interested in evaluating:

- **Effectiveness:** were the two goals of exploration achieved?
- **Style:** was exploration *broad*, exploring many combinations of dimensions, or *deep*, emphasising word inspection and navigation within dimension pairs?
- **Usability:** do users find the system usable?

We conducted an experiment to assess these questions, comparing our interface with the following two alternatives. Firstly, a *plain scatterplot* system replaces the cluster visualisation with a scatterplot that users can pan and zoom – a conventional clutter management strategy (Fig. 6a). Secondly, a *heatmap* system uses the KDE heatmap as the primary display (Fig. 6b). To view individual points, the user selects an area of the plot, within which individual points are rendered. When the selection is made, the plot zooms into the selection, and individual points are rendered which can be inspected using tooltips. Both systems retain the navigation matrix, but lose the graphical history and word cloud, leaving tooltips as the method for word inspection.

### 4.1 Experiment Design and Procedure

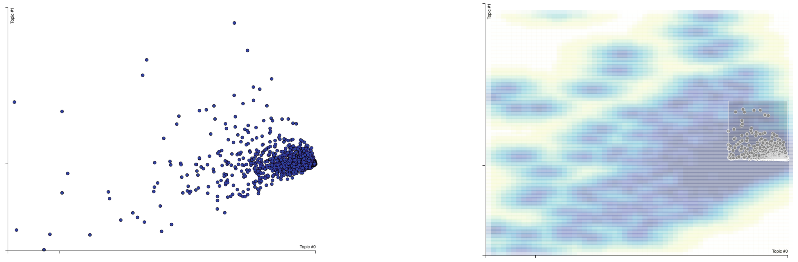
Twelve Cambridge University undergraduates with no prior exposure to LSA were recruited via convenience sampling. The experiment was a within-subjects comparison of the three systems. Participants were briefed on LSA and on the operation of each interface. For each of the three interfaces, participants carried out an experimental task, then completed two usability questionnaires.

In each task, the participant was instructed to (i) assign a meaning to groups of words which they found to be related, and (ii) assign a meaning to each of the dimensions. Three disjoint datasets of 30,000 words were sampled from the corpus. The design was counterbalanced: each dataset was assigned to each interface in equal representation across participants. The 3 systems were presented to participants in different orders, each of the 6 possible orders being assigned exactly twice. These measures mitigated learning effects and order effects.

We recorded the number of meanings offered by the user, counting at most one assigned meaning per word group/dimension, even if the participant offered multiple interpretations. Participants were free to continue exploration as long as they desired. General usability ratings were obtained using IBM's Post-Study System Usability Questionnaire (PSSUQ), while IBM's After-Scenario Questionnaire (ASQ) was used to measure task-specific usability [21]. Both use a 7-point scale with lower values reflecting superior usability.

### 4.2 Results

We refer to our interactive Cluster diagram as C, the Heatmap alternative as H, and the plain Scatterplot alternative as S. All post-hoc tests were subjected to Bonferroni correction.



(a) Plain scatterplot with pan/zoom.

(b) Heatmap showing selected points.

**Fig. 6.** Experimental alternative interfaces

*Assignment of Meaning*: Participants assigned meaning to significantly more word groups using C (average of 7.92 groups) versus H (5.33 word groups, $p = 0.037$) and S (4.25, $p = 0.038$). (Planned contrasts after one-way repeated measures ANOVA yielded $F(2, 22) = 5.162$, $p = 0.019$). A significant difference was not found in the number of meaning assignments for dimensions.

*Style of Exploration*: We studied how often users switched the dimensions displayed using the heatmap matrix. Participants switched dimensions several times in S, but less frequently when using C. In contrast, participants inspected a far greater number of words with C than with either alternative. C therefore promoted a more depth-first style of exploration due to the ease of navigating the hierarchy, facilitating model interpretation grounded in specific words. Concretely: a significant Friedman's test was followed with Wilcoxon signed rank tests. Users switched dimensions more often with S ($p = 0.037$). With a similar analysis, the number of words inspected was significantly different ($p = 0.028$). The average number of words explored when using C was 1517 ($p = 0.010$), as compared to 479 with H and 772 with S ($p = 0.050$).

*Usability*: Users found C more usable than either alternative. C significantly improved the users' exploration effectiveness in terms of the number of groups of related words found. This was expected, as the word cloud allows more words to be inspected simultaneously, and clusters encapsulate the semantics of a given word group. Concretely: The PSSUQ score for C (average 1.86) was significantly better than H (average 4.08, $p = 0.002$) or S (average 3.01, $p = 0.015$) (Wilcoxon signed-rank tests following significant Friedman's test ($p = 0.001$)). The differences in task-specific ASQ ratings for the *dimension interpretation* task were significant ($p = 0.002$). C was rated better than both S (mean difference $-1.389$, $p = 0.02$), and H (mean difference $-1.528$, $p = 0.001$). For the *word group* task we observed similar, but non-significant mean differences.

## 5 Conclusion

Latent semantic spaces are a valuable tool for the analysis of large text corpora. However, interpreting latent semantic spaces is difficult, and visual scalability is a major design challenge, as is accessibility for non-experts.

Our novel interface uses a hierarchical clustering approach to clutter reduction, allowing users to gain an overview of semantic structure in the corpus. The cluster diagram can be combined with summary distributions arranged in a heatmap matrix. A user study showed that the usability of our interactive diagram was significantly superior to alternatives based on either plain scatterplots or heatmaps alone. Moreover, the hierarchical cluster diagram facilitated the identification and assignment of meaning to more word groups.

## References

1. Landauer, T.K., Foltz, P.W., Laham, D.: An introduction to latent semantic analysis. Discourse Process. **25**(October), 259–284 (1998)
2. Turney, P.D., Pantel, P.: From frequency to meaning: vector space models of semantics. J. Artif. Intell. Res. **37**, 141–188 (2010)
3. Landauer, T.K., Laham, D., Rehder, B., Schreiner, M.E.: How well can passage meaning be derived without using word order? A comparison of latent semantic analysis and humans. In: Proceedings of the Cognitive Science Society, pp. 412–417 (1997)
4. Landauer, T.K., Laham, D., Derr, M.: From paragraph to graph: latent semantic analysis for information visualization. Proc. NAS USA **101**, 5214–5219 (2004)
5. Habernal, I., Brychcín, T.: Semantic spaces for sentiment analysis. In: Habernal, I., Matoušek, V. (eds.) TSD 2013. LNCS (LNAI), vol. 8082, pp. 484–491. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40585-3_61
6. Elmqvist, N., Dragicevic, P., Fekete, J.D.: Rolling the dice: multidimensional visual exploration using scatterplot matrix navigation. IEEE TVCG **14**(6), 1141–1148 (2008)
7. Fua, Y.H., Ward, M.O., Rundensteiner, E.A.: Hierarchical parallel coordinates for exploration of large datasets. In: Proceedings of the Conference on Visualization, pp. 43–50 (1999)
8. Carr, D.B., Littlefield, R.J., Nicholson, W.L., Littlefield, J.S.: Scattterplot matrix techniques for large N. J. Am. Stat. Assoc. **82**(398), 424–436 (1987)
9. Zhu, W., Chen, C.: Storylines: visual exploration and analysis in latent semantic spaces. Comput. Graph. **31**, 338–349 (2007)
10. Ellis, G., Dix, A.: A taxonomy of clutter reduction for information visualisation. IEEE Trans. Vis. Comput. Graph. **13**(6), 1216–1223 (2007)
11. Elmqvist, N., Fekete, J.D.: Hierarchical aggregation for information visualization: overview, techniques, and design guidelines. IEEE Trans. Vis. Comput. Graph. **16**(3), 439–454 (2010)
12. Shaoul, C., Westbury, C.: The Westbury Lab Wikipedia Corpus. University of Alberta, Edmonton (2010)
13. Rehurek, R., Sojka, P.: Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, Valletta, Malta, ELRA, pp. 45–50, May 2010
14. Baker, K.: Singular value decomposition tutorial. Ohio State Univ. **24**, 16–23 (2005)
15. Landauer, T.K., McNamara, D.S., Dennis, S., Kintsch, W.: Handbook of Latent Semantic Analysis (2007)
16. Salvador, S., Chan, P.: Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In: 2012 IEEE 24th International Conference on Tools with Artificial Intelligence, pp. 576–584 (2004)
17. Pirolli, P., Card, S.: Information foraging. Psychol. Rev. **106**(4), 643 (1999)
18. Wang, M.P., Jones, M.C.: Comparison of smoothing parameterization in bivariate kernel density estimation. J. Am. Stat. Assoc. **88**(422), 520–528 (1993)
19. Kurlander, D., Feiner, S.: Editable graphical histories. In: IEEE Visual Languages, pp. 127–134 (1988)
20. Cockburn, A., Karlson, A., Bederson, B.B.: A review of overview+detail, zooming, and focus+context interfaces. ACM Comput. Surv. **41**, 1–31 (2008)
21. Lewis, J.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. IJHCI **7**(1), 57–78 (1995)

# How Users Transform Node-Link Diagrams to Matrices and Vice Versa

Simone Kriglstein(✉), Margit Pohl, and Johanna Doppler Haider

Vienna University of Technology, Vienna, Austria
simone.kriglstein@tuwien.ac.at, {margit,johanna.haider}@igw.tuwien.ac.at

**Abstract.** A combination of node-link diagram and matrix seems to be beneficial since their respective strengths and weaknesses complement each other. However, users have to read both representations in different ways and switch between these representation styles. We conducted a user study to understand how users transform a node-link diagram to a matrix representation and vice versa. For this purpose we let participants draw node-link diagrams and matrices. The drawings were analyzed to identify strategies how user convert one visualization into the other one.

**Keywords:** Information visualization · Node-link diagram · Matrix

## 1 Introduction

The analysis of graphs is essential for many areas, e.g., social network analysis, biology, or medicine. The traditional approach is to visualize graphs as node-link diagrams. This kind of visualization is easily understandable for many users. Nevertheless, there are some drawbacks related to node-link diagrams. Even relatively small graphs result in large node-link diagrams and dense graphs lead to clutter so that nodes and links can be occluded. As an alternative, matrix visualizations have been suggested because they use space more efficiently and are free of clutter. The disadvantage of this kind of visualization is that inexperienced users have difficulties in interpreting matrices [14] and that the detection of relationships is not as intuitive as in node-link diagrams.

In this paper we try to clarify how users translate node-link diagrams into matrices and vice versa which can inform the design of hybrid visualizations. When using such hybrid visualizations users have to relate the information gained from one visualization to the information gained from the other one. We hope that this study also contributes to a more detailed understanding of the problem why matrices are not as easily understandable as node-link diagrams. We want to analyze what conceptual problems may occur when people engage in such activities. To achieve this we asked the participants to look at either node-link diagrams or matrices and draw a sketch of the other visualization based on the same data. We then analyzed the errors which occurred and, more generally, the strategies the users adopted to relate both visualizations with each other.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 526–534, 2018.
https://doi.org/10.1007/978-3-319-91376-6_48

## 2 Related Work

There is a considerable amount of research about the advantages and disadvantages of node-link diagrams and matrices and how to design these visualizations. Ghoniem et al. [5] used simple, generic tasks and found out that matrices are especially useful for larger, denser networks. Graphs are suited for path related tasks (cf. [9]). Henry and Fekete [7] developed MatLink, a hybrid tool consisting of node-link diagram and matrix. To support path-related tasks MatLink shows graphical curved links of the connecting nodes in lines and columns. In an evaluation, they showed that MatLink is superior to either node-link diagram or matrix. Alper et al. [1] compared weighted node-link diagrams and matrices and also found that matrix representations in general are more efficient than node-link diagrams. In general, those studies used generic and fairly simple tasks (e.g., find a node or identify a path between two nodes). Some visualization tools combine matrices and node-link diagrams, such as the MatrixExplorer [6], where a node-link diagram is juxtaposed with an adjacency matrix. Another combination of node-link diagrams and matrices is implemented in NodeTrix [8], where dense sub-networks are depicted as adjacency matrices and the sparse parts of the network are drawn as links that connect the matrices. Liu and Shen [11] show that square matrices lead to improved pattern recognition in comparison to triangular matrices and propose a compact juxtaposition of many matrices. There is also some research proposing to use multiple links between nodes. Node-link diagrams using multivariate edges, so called multiple threads, were presented in Ko et al. [10]. They argue that there are many application areas for such multiple edges, but state that such multiple links are sometimes difficult to distinguish and that an appropriate design is important. Beck et al. [3] also discuss the usage of different multiple links.

## 3 Empirical Evaluation

The study aimed to address the following research questions:

- **RQ1 - Correctness and Efficiency:** Do participants convert node-link diagrams to matrix or vice versa more correctly? Does transforming node-link diagrams to matrices or vice versa yield faster completion times?
- **RQ2 - Interpretation:** How do participants convert node-link diagrams to matrices and vice versa? How do they transform the connections between those two representations? How do they transform the nodes from a node-link diagram to a matrix and vice versa?

Since hand drawn sketches are an effective method that allow for creative freedom and help to express, develop, and communicate concepts with a low entry barrier (cf. [4,15,16]), we decided to let participants draw node-link diagrams and matrices to avoid any restrictions or influences resulting from any software.

*Materials.* For the evaluation two test cases were generated. For each test case, node-link diagram/matrix examples of the following four different types were prepared: *Type 1* - 1 connection type with 5 nodes, *Type 2* - 3 different connection types with 5 nodes, *Type 3* - 3 different connection types with 10 nodes, and *Type 4* - 3 different connection types with direction information and 8 nodes.

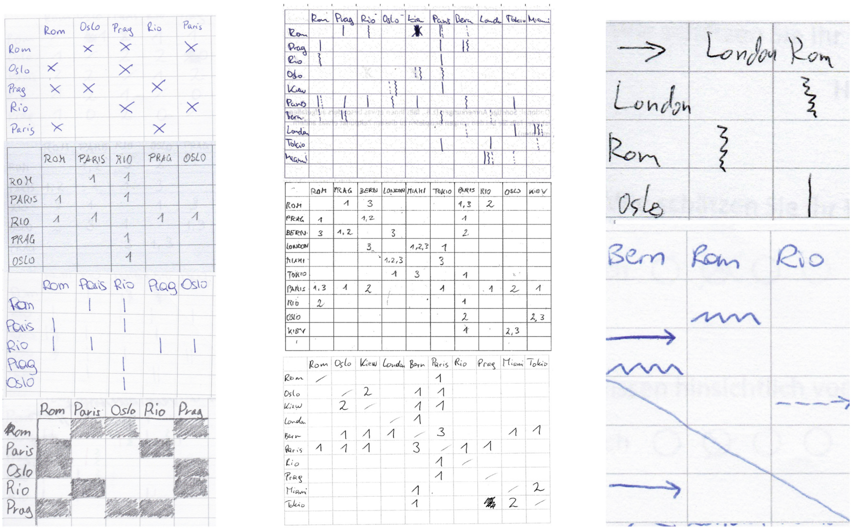
The goal of *Type 1* is to verify if participants understand the relationship between a node-link diagram and its corresponding matrix and vice versa. In order to avoid errors which may result from the complexity imposed by a large number of nodes and connections, we decided to restrict the examples to five nodes. For *Type 2* we increased the complexity by increasing the number of connection types. For better differentiation of the three connection types, different line types were used. The number of nodes stayed the same as in *Type 1* to ensure that the connection types and not the number of nodes influence the result. In contrast to *Type 2*, we doubled the number of nodes for *Type 3* to investigate if the number of nodes has any influence on the outcome. Finally, the focus of *Type 4* was on the direction information of the connection types and how participants will transform this information between the two representations. We decided to use city names and the connection types represent different modes of transport.

*Sample.* We recruited 12 Computer Science students (6 males and 6 females with an average age of 25.16) who have at least basic knowledge with matrix and node-link visualizations. No compensation was offered.

*Procedure.* The experimental session was conducted in a quiet seminar room. The session started with a short introduction and participants had the possibility to ask questions to clarify any issues with the study. Next, the participants got the two test cases as discussed above. To counteract learning effects the participants were divided into two groups. The first group of participants got the four matrices from one test case and were asked to draw the corresponding node-link diagrams. After they finished, they got the four node-link diagrams from the other test case and were asked to draw the corresponding matrices. For the second group, the order was reversed. The average duration for drawing all node-link and matrix examples was about 47 min (min. 37 min and max. 60 min). Lastly, the participants were asked if they found it laborious to convert matrices into the corresponding node-link diagram or/and vice versa.

## 4 Results

*RQ1 - Correctness and Efficiency.* First, we assessed if participants converted node-link diagrams to matrices and vice versa correctly. In total 32 errors were made by 11 out of 12 participants. One participant did not make a single mistake. The analysis of errors showed only minimal differences if participants converted node-link diagrams to matrices (17 errors) or vice versa (15 errors). However, an interesting observation was that of the 11 participants who made errors, four participants made the errors only when converting matrices to node-link diagrams,



(a) (b) (c)

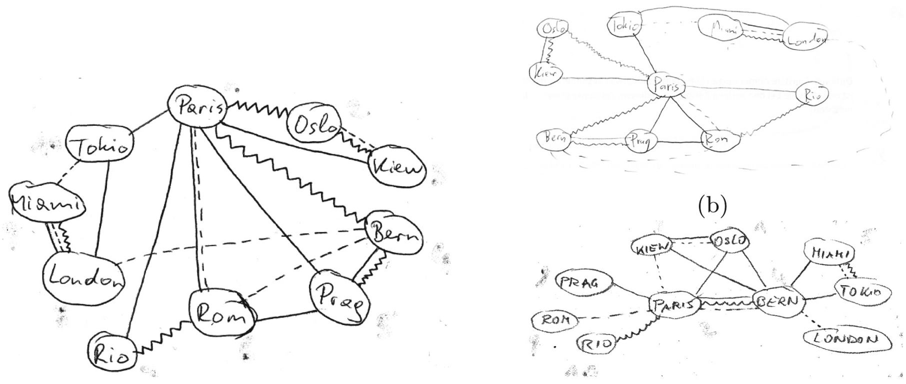
**Fig. 1.** Matrix examples for (a) *Type 1*, (b) *Type 2* and *3*, and (c) *Type 4*.

five participants only when converting node-link diagrams to matrices, and only two participants made errors in both direction. Most of the errors were made with the examples of *Type 4* ($n = 19$), followed by *Type 3* ($n = 9$) and *Type 2* ($n = 3$). Only one person made an error with an example of *Type 1*. This person doubled the connections when converting the matrix to a node-link diagram. Asked about the reason, the participant noted that the matrix included each entry twice (due to symmetry reason). Although 13 of the 32 errors (made by 4 participants) were semantic errors (e.g., only the number of connection types was added in the matrix without making a difference between them), most errors ($n = 19$) which were made by 11 participants were syntax errors (e.g., connections between nodes were missed). Since syntax errors were mainly mistakes due to inadvertences but semantic errors resulted from misinterpretations, most of the participants (8 out of 12) transform node-link diagrams to matrices and vice versa correctly. We could observe that the participants took slightly longer to convert the node-link diagrams to matrices (27 min for all four examples) than for the other way round (21 min).

*RQ2 - Interpretation.* We observed that five participants preferred to convert matrices to node-link diagrams and seven participants favored to convert node-link diagrams into matrices. Participants noted the following reasons for their preference to convert node-link diagrams to matrices: the raster layout from the matrix (2 statements) and the clear arrangement of the matrix (3). Participants who favored to convert matrices to node-link diagrams noted that: it made more fun (1) and it was easier to understand (2) than drawing matrices. Furthermore, one participant mentioned that nodes and edges needed to be drawn twice for matrices which is not necessary for node-link diagrams. However, on average converting matrices to node-link diagrams and converting node-link diagrams to matrices was found equally laborious in both cases. Next, we analyzed how participants transformed node-link diagrams to matrices and vice versa.

Node-Link Diagrams to Matrices: Most of the participants (7 out of 12) drew a × symbol in the matrix examples of *Type 1* to represent connections between the nodes. A line and number *1* as a symbol were used by two participants each. One person grayed out the cell in the matrix. Figure 1(a) shows four examples how the connections were drawn in the matrix. For the matrix examples of *Type 2–4*, we observed that most of the participants used the same line types in the matrix as used in the node-link diagram (*Type 2*: 8 participants, *Type 3*: 9, and *Type 4*: 9). The image at the top in Fig. 1(b) shows an example how participants represented different connection types with lines in the matrix. However, it was interesting to observe that participants did not pay much attention to the order of the line types. For only about 50% of the examples the order of the line types in the matrices corresponded to the order of the line types in the node-link diagrams. Four participants used numbers instead of the different line types. Two participants only used the sum of the number of connections between nodes. Two other participants assigned the different connection types to the numbers *1*, *2*, and *3* and wrote the corresponding numbers in the cells (see Fig. 1(b)). These two participants changed the presentation style for examples of *Type 3* and *Type 4*. One participant used the different line types instead of the numbers. The second participant used different symbols for the example of *Type 4*. The four participants, who used numbers, converted the node-link diagrams to matrices first before they converted the matrices to node-link diagrams. The analysis of the examples of *Type 4* showed that 10 participants give the reading direction of the connections along the left side of the matrices. Only two persons drew the direction information of the connections directly within the cells of the matrices. Again, we could observe that these two participants drew the matrix examples first. Figure 1(c) shows excerpts of examples how participants drew the direction information in the matrix. The most used strategy to transfer the nodes from the node-link diagrams to matrices was clockwise or counterclockwise if the arrangement of nodes in the node-link diagram exhibited a circular structure. However, if the arrangement of nodes in the node-link diagram did not resemble a circular layout, other strategies such as left-to-right or top-to-bottom, were used. Interestingly, the structure of the edges did not seem to play an important role for how the participants converted the node-link diagrams to matrices. A similar phenomenon was also observed by Ballweg et al. [2] who noticed that edges did not play an important role for the similarity perception of graphs.

Matrices to Node-Link Diagrams: We could observe that in 57% of the examples participants arranged nodes in a star-like fashion with one node being drawn in the center (see Fig. 2(b) for an example) if they went through the nodes from left-to-right. If they went through the nodes clockwise or counterclockwise, the nodes were arranged circular in 66% of the cases (see, e.g., Fig. 2(a)). In total,



(a)

(b)

(c)

**Fig. 2.** Examples for (a) circular layout, (b) star layout, and (c) focus on edges.

the circular layout was used slightly more often (41.6% of the examples) than the star layout (33.3%). Only for the examples of *Type 3* and *Type 4*, participants considered the edges when arranging the nodes to minimize edge crossings (*Type 3*: 33% of the examples; *Type 4*: 41%). Figure 2(c) shows an example. For the remaining examples star layout and circular layout were used. Consideration of edges was also given as reason why participants changed the arrangement of nodes, which they used for examples of *Type 1* and *2*, for the examples of *Type 3* and *4*. Similar to the converting node-link diagrams to matrices, participants did not pay much attention to the order of the different connection types if they transformed them from matrices to node-link diagrams. The analysis of the examples of *Type 4* showed no clear preference for drawing the directions of the edges in node-link diagrams. Four participants doubled the lines to represent both directions, three participants used a single line with an arrow head for each direction, and five participants drew a single line with arrow heads for both directions.

## 5 Discussion

*Node-Link Diagram to Matrix Conversion.* The analysis of the matrix sketches showed us that participants who drew the matrices first, converted the node-link diagrams to matrices differently. For example, numbers were used for matrices instead of different line types to represent the connection types between nodes. In return, there were no differences in converting matrices to node-link diagrams independently if participants first converted node-link diagrams to matrices or vice versa. This led us to the assumption that how node-link diagrams should look like is unequivocal whereas the representation of matrices allows different options (e.g., numbers vs. symbols).

*Order of Line Types.* The results showed us that most of the participants made differences between different representations of the connections but they did not pay much attention to the order of the different connection types when converting one visualization into the other. Although it seems that participants paid more attention in case of the smaller node-link diagrams with five nodes than in case for node-link diagrams with the double number of nodes. A reason could be that no information about the order of the connection types was provided and hence they focused only on the obvious differences between the connection types.

*Arrangement of Nodes.* The results showed us that the *Law of Good Gestalt* (cf. [12]) seems to play a role for how the participants arranged the matrices and node-link diagrams. In this context, an interesting observation is that the clockwise or counterclockwise arrangement of nodes seems to be dependent on if they form a simple pattern. The *Law of Good Gestalt* could also be the reason that most of the participants preferred a circular and star layout for node-link diagrams since these layouts can be seen as simple, orderly, coherent, and regular. In addition, arranging nodes in an orderly fashion allowed the participants to preserve the mental model of the graph more easily.

*Importance of Edges.* We could observe that the number of edges as well as the structure of the edges does not seems to play an important role how the participants converted the node-link diagrams to matrices. This was interesting since the arrangement of nodes based on the edges can be helpful to see patterns, e.g., highly connected nodes. A possible reason for ignoring the edges can be that no analysis of the node-link diagram or the corresponding matrix was required. In contrast to converting node-link diagram to matrix, it seems that graph drawing aesthetics [13] gain in importance (e.g., minimization of edge crossings) if the number of nodes increased. This corresponds with the answers of participants preferring matrices; the raster layout from the matrix allowed a clear arrangement. This is an advantage over node-link diagrams which can be easily confusing because of possible edge crossings.

## 6 Conclusion

In this paper, we presented a user study conducted to get a more detailed understanding of the conceptual problems which may occur when people have to convert node-link diagrams to matrices and vice versa. The analysis of the sketches of the participants showed us, that the representation of matrices allowed more creative opportunity to present the connections than the node-link diagrams. Another interesting observation was that the participants did not pay much attention to the order of the connection types after they converted them from a node-link diagram to matrix or vice versa.

As future work, we will address the question if edges and the order of the connection types will gain in importance when participants have to analyze and interpret the node-link diagrams and matrices. In the next phase of this research we will also investigate how interaction strategies will influence users' interpretation of the data if they have to relate the information gained from one visualization to the information gained from the other visualization.

**Acknowledgments.** This work was supported by CVAST, funded by the Austrian Federal Ministry of Science, Research, and Economy in the exceptional Laura Bassi Centres of Excellence initiative (#822746).

## References

1. Alper, B., Bach, B., Henry Riche, N., Isenberg, T., Fekete, J.D.: Weighted graph comparison techniques for brain connectivity analysis. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM (2013)
2. Ballweg, K., Pohl, M., Wallner, G., von Landesberger, T.: Visual similarity perception of directed acyclic graphs: a study on influencing factors. In: Frati, F., Ma, K.-L. (eds.) GD 2017. LNCS, vol. 10692, pp. 241–255. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73915-1_20
3. Beck, F., Burch, M., Diehl, S., Weiskopf, D.: The state of the art in visualizing dynamic graphs. In: EuroVis State-of-The-Art Report. The Eurographics Association (2014)
4. Buxton, B.: Sketching User Experiences: Getting the Design Right and the Right Design. Morgan Kaufmann Publishers Inc., Burlington (2007)
5. Ghoniem, M., Fekete, J.D., Castagliola, P.: A comparison of the readability of graphs using node-link and matrix-based representations. In: Proceedings of the IEEE Symposium on Information Visualization. IEEE (2004)
6. Henry, N., Fekete, J.D.: MatrixExplorer: a dual-representation system to explore social networks. TVCG **12**(5) (2006)
7. Henry, N., Fekete, J.-D.: MatLink: enhanced matrix visualization for analyzing social networks. In: Baranauskas, C., Palanque, P., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4663, pp. 288–302. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74800-7_24
8. Henry, N., Fekete, J.D., McGuffin, M.J.: NodeTrix: a hybrid visualization of social networks. TVCG **13**(6) (2007)
9. Keller, R., Eckert, C.M., Clarkson, P.J.: Matrices or node-link diagrams: which visual representation is better for visualising connectivity models? Inf. Vis. **5**(1) (2006)
10. Ko, S., Afzal, S., Walton, S., Yang, Y., Chae, J., Malik, A., Jang, Y., Chen, M., Ebert, D.: Analyzing high-dimensional multivariate network links with integrated anomaly detection, highlighting and exploration. In: Proceedings of the IEEE Conference on Visual Analytics Science and Technology. IEEE (2014)
11. Liu, X., Shen, H.W.: The effects of representation and juxtaposition on graphical perception of matrix visualization. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM (2015)
12. Metzger, W.: Laws of Seeing. The MIT Press, Cambridge (2006)
13. Purchase, H.C.: Metrics for graph drawing aesthetics. JVLC **13**(5) (2002)

14. Seidler, P., Haider, J., Kodagoda, N., Wong, B.L.W., Pohl, M., Adderley, R.: Design for intelligence analysis of complex systems: evolution of criminal networks. In: Proceedings of the European Intelligence and Security Informatics Conference. IEEE (2016)
15. Tversky, B., Suwa, M.: Thinking with Sketches. Oxford University Press, Oxford (2009)
16. Walny, J., Huron, S., Carpendale, S.: An Exploratory Study of Data Sketching for Visual Representation. In: Computer Graphics Forum (2015)

# Peirce and Existential Graphs

# Peirce and Proof: A View from the Trees

Dave Beisecker(✉)

University of Nevada, Las Vegas, Las Vegas, NV 89154, USA
beiseckd@unlv.nevada.edu

**Abstract.** Using the proof of Peirce's Law [$\{(x \to y) \to x\} \to x$] as an example, I show how bilateral tableau systems (or "2-sided trees") are not only more economical than rival systems of logical proof, they also better reflect the reasoning Peirce actually gives for securing the law's acceptance as an axiom. Moreover, bilateral proof trees are readily adapted to Peirce's own graphical notation, producing a proof system in that notation that is even more efficient and easier to learn than Peirce's system of permissions. This is in part due to the fact that Peirce's graphical notation is similarly bilateral. In effect bilateral proof trees in Peirce's notation can be understood as representing the space of outcomes for a game very much like what Peirce envisions as his endopereutic, and they embody insights of certain expressions of the pragmatic maxim that Peirce offers around 1905. Taken together, this suggests to me that Peirce would have embraced such a system of logic, and so I find it especially unfortunate that he was evidently unaware of Lewis Carroll's pioneering efforts to develop tree-like proof systems to solve logical puzzles with multiliteral sorites.

**Keywords:** C. S. Peirce · Proof · Tableau · Trees · Graphs · Lewis Carroll

Let's begin by considering the fifth (and final) axiom that Peirce presents in his 1885 "On the Algebra of Logic" [10]:

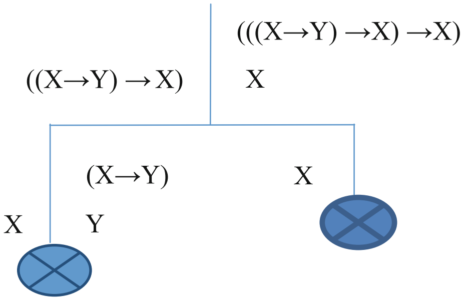
> A fifth icon is required for the principle of excluded middle and other propositions connected with it. One of the simplest formulae of this kind is: $\{(x \to y) \to x\} \to x$.

This schema has come to be known as "Peirce's Law," and while he notes that the form hardly appears self-evident enough to be axiomatic, what intrigues me is the form of justification that Peirce subsequently gives to secure its acceptance. He continues:

> This is hardly axiomatical. That it is true appears as follows. It can only be false by the final consequent x being false while its antecedent $(x \to y) \to x$ is true. If this is true, either its consequent, x, is true, when the whole formula would be true, or its antecedent $x \to y$ is false. But in the last case the antecedent of $x \to y$, that is x, must be true (CP 3.384).

Note how Peirce begins this justification by working out the consequences of *denying* the law – of holding it false – and then showing that the subsequent pattern of affirmations and denials that that denial carries in its train inevitably leads to a contradiction or collision of commitments – of both affirming and denying one and the same proposition. Peirce's reasoning is displayed virtually line-by-line in the corresponding bilateral (2-sided) diagrammatic tree proof of the law:

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 537–548, 2018.
https://doi.org/10.1007/978-3-319-91376-6_49



The tree (or Tableau) method of proof works by developing the consequences of affirming some claim or claims (starting with premises) and also the consequences of denying claims (beginning with conclusion(s)), with the aim of exposing incompatibilities between patterns of affirmations and denials. In the "one-sided" trees more commonly found in logic texts, the consequences of denying a claim is represented as the consequences of affirming that claim's negation. One-sided (or "Smullyan") trees thus have a symbolic element (the negation *sign*) occupying a role that could instead be played by a visual or diagrammatic feature (alternating *sides* of a path). Intuitively, one can see why a two-sided tree system would be commensurate with natural deduction systems deploying the standard Fitch-style elimination and introduction rules for the various sentential operators (See Fig. 1). The rules for developing affirmations of certain contents obviously correspond to elimination rules, while the rules for developing denials of contents correspond to the introduction rules – except (as explicitly suggested by one-sided trees) that they work in the "opposite" direction. In essence, the right hand (or negation decomposition) rules in Tableau, serve to "tollens" the introduction rules.[1]

One occasionally encounters a sentiment to the effect that tree diagrams such as the one above are not truly "proofs" at all. Rather, the method of trees is just an especially efficient trick or technique for navigating truth tables - which can happily be extended to first-order and modal logics - with the aim of honing in on possible counterexamples. By contrast, "bona-fide" proof systems, such as axiom or natural deduction systems, are supposed to mimic more closely step-by-step patterns of natural reasoning in language. I think the fact that this tree proof parallels so closely the argument that Peirce actually gives for the axiom should put to rest any such sentiment that such tree or tableau systems are not "really" systems of proof at all.

Indeed, it pays to contrast this tree proof with the corresponding derivation of the same law using a standard system of natural-deduction using Gentzen-style rules of inference.

---

[1] With that in mind, one can readily discern what the Tableau rules for Belnap's infamous "Tonk" operator would be, even though it would be impossible to read such rules off of a truth table: namely the left-hand rule of a conjunction, paired with the right-hand rule of a disjunction.

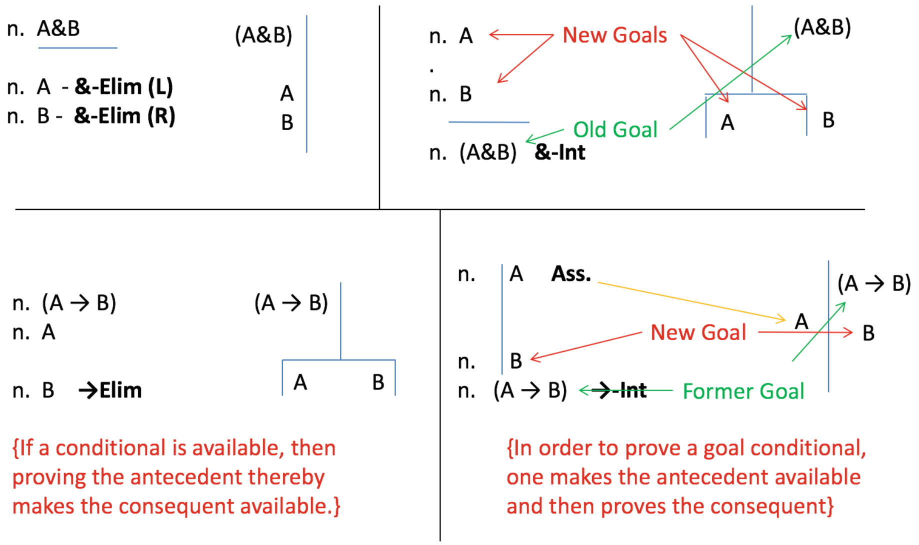


**Fig. 1.** Correspondences between ND inference rules and Tableau decomposition rules. Elimination rules correspond to left-side decomposition rules, while introduction rules correspond to right-side decomposition rules.

Such a proof is not nearly as straightforward, containing as it does, two refutations (or applications of negation elimination) embedded within a conditional proof:

| | | |
|---|---|---|
| 1. | (X→Y) →X | Assumption (→ Int.) |
| 2. | ~X | Assumption (~ Elim.) |
| 3. | X | Assumption (→ Int.) |
| 4. | ~Y | Assumption (~ Elim.) |
| 5. | ⊥ | ⊥ Int.; 2, 3 |
| 6. | Y | ~ Elim.; 4-5 |
| 7. | X→Y | → Int.; 3-6 |
| 8. | X | → Elim.; 1, 7 |
| 9. | ⊥ | ⊥ Int.; 2, 8 |
| 10. | X | ~ Elim.; 2-9 |
| 11. | ((X→Y) →X) →X | → Int.; 1-10 |

The construction of this derivation requires a bit of ingenuity that isn't needed for the corresponding tree proof, and despite the system's bearing the label "natural deduction", it is definitely *not* how one would naturally reason toward such a claim! Indeed, the proof of Peirce's Law is sometimes used as a test to separate those who have acquired genuine mastery of the construction of such derivations from the "merely competent." The chief difficulty (and source of frustration) with the construction of natural-deduction derivations stems from the system's relative open-endedness. It's not always clear just how one should begin, or how to go about pursuing subsequent subgoals. The specific order of applicable rule applications matters. Moreover, it takes a certain measure of intuition or insight to “see” when one should back out of a direct strategy for deriving a formula in favor of a different strategy, including that of pursuing a goal formula indirectly by refuting its negation.[2] None of these are issues for tree proofs. There are no indirect strategies, and for any compound formula in a tree proof, there is only one applicable rule to decompose that formula. Furthermore, though some proofs may be more unwieldy than others on account of their having more branches than is necessary, the specific order of rule applications doesn’t matter in the end. As a result, it takes far less technique and practice to master tree systems, and the method requires virtually no intuition whatsoever.[3]

But what of Peirce’s own graphical notation and method of proof – his “moving pictures of thought”? Peirce’s graphical system of logic certainly has its fans [12, 14], and it does possess an impressive economy (largely borne of the fact that its notation collapses distinctions between separate but equivalent formulas in more familiar symbolic notations). That makes proofs in the system relatively shorter than in more conventional systems. For instance, Fig. 2 depicts a proof of Peirce’s Law that is just 8 steps long.[4]

There is even an evident bilaterality built into Peirce’s diagrams that begins to resemble the bilaterality of two-sided trees. We begin with one side of a piece of paper, which represents the sheet of assertion (the “recto”). But that sheet of paper is to be understood as having a reverse side (the “verso”), upon which one lists claims to be denied. The representation of a negation is to be understood then by drawing a circle around a denied claim. Peirce tells us that this is like “cutting” a hole in the sheet around a denied claim and then “flipping it” around so that it appears upon the sheet of assertion.

> Should the Graphist desire to negative a Graph, he must scribe it on the verso, and then, before delivery to the interpreter, must make an incision, called a Cut, through the Sheet all the way round the Graph-instance to be denied, and must then turn over the excised piece so as to expose its rougher surface carrying the negativized Graph-instance (CP 4.556).

In short, Peirce intends his logical diagrams to be understood as bilateral (or two-sided) in the most literal sense of the term! Cuts on a sheet of either assertion or denial give us a glimpse of what appears on “the other side.” In his system of existential graphs, proofs are transformations of graphs (“moving pictures of thought”) according to rules or permissions, which in effect depend upon which side of the sheet of paper (front or reverse) a claim is inscribed. Roughly, elements may be added to graphs on the reverse side, and deleted from graphs on the front (“recto”). As Peirce explains, these operations are akin to stacking conditions into the antecedent of conditionals and removing conditions from their consequences (CP 4.564).

---

[2] Chapters 5–6 of Wilfried Sieg’s *Logic and Proofs* [13] contains perhaps the best discussion of strategies for going about constructing natural-deduction derivations using standard introduction and elimination rules.

[3] Axiomatic systems, of course, fare even worse by comparison. For instance, an axiomatic derivation of Peirce’s law using Mendelson’s three-axioms (which of course do not include Peirce’s Law) takes over 30 steps, and requires a rather ingenious selection of initial forms of the axiom schema.

[4] One can actually shorten this proof to seven steps by combining the insertions in steps 5 and 8. I find the proof in Fig. 2 to be more perspicuous in that it shows why the content of the consequent in the most deeply embedded conditional (the Y that is the second insertion in step 8) is immaterial to the overall truth of the schema.

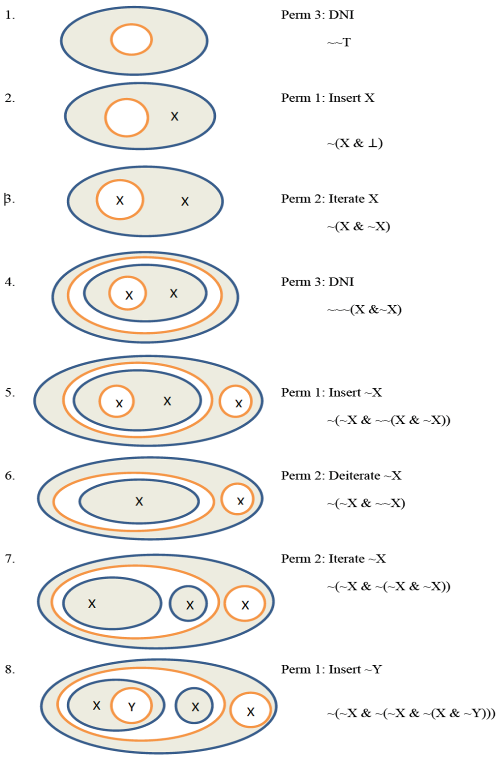


**Fig. 2.** Proof of Peirce's Law using Peirce's system of permissions

Nevertheless, as a proof system, Peirce's system of permissions is still not as efficient as bilateral proof trees. Nor does it resemble the line of reasoning Peirce actually deploys to justify Peirce's Law. Unlike tree proofs, Peirce's system lacks a ready means of illustrating collisions of commitments or incompatibilities, and while Peirce's permissions are sensitive to the side on which a given graph element is, as it were, inscribed, the system displays the proof only from the side of assertion. In that respect, it is like more conventional axiomatic and natural deduction proof systems, and it suffers from the same open-endedness that those more familiar systems bear in comparison to tree systems. A certain intuition or creativity is required on the part of Peirce's graphist to craft a desired proof, especially with respect to the free choices of elements to insert into a "verso" portion of a graph. By contrast, bilateral (two-sided) tree proofs depict the sides of assertion and denial on the same page at once as different sides of a path, and then they display on different branches the various possible consequences of a given pattern of commitments of assertions and denials.

But note that one can readily come up with a bilateral tree system using Peirce's own notation (at least for his alpha graphs). Since the notation for graphs are simply composed of concatenations of simpler graphs (which may be understood as conjunction) and cuts (which may be understood as negation), all one needs to do is to adapt graph decomposition rules for concatenation and the cut that mimic the regular bilateral tree rules for conjunction and negation. Concatenations of graphs on the left are simply pulled apart and then placed separately on the left, while concatenations of graphs on the right require one to split a branch and place separate concatenends on the right of each resulting branch. The rule for cut is even more straightforward; for a cut that is either on the left or the right, simply erase the cut and place the remainder on the opposite side. As in other bilateral tree systems, branches close when one and the same formula (or graph) appears on both the left and the right sides of a path. The tree proof for Peirce's law using this system is displayed in Fig. 3.

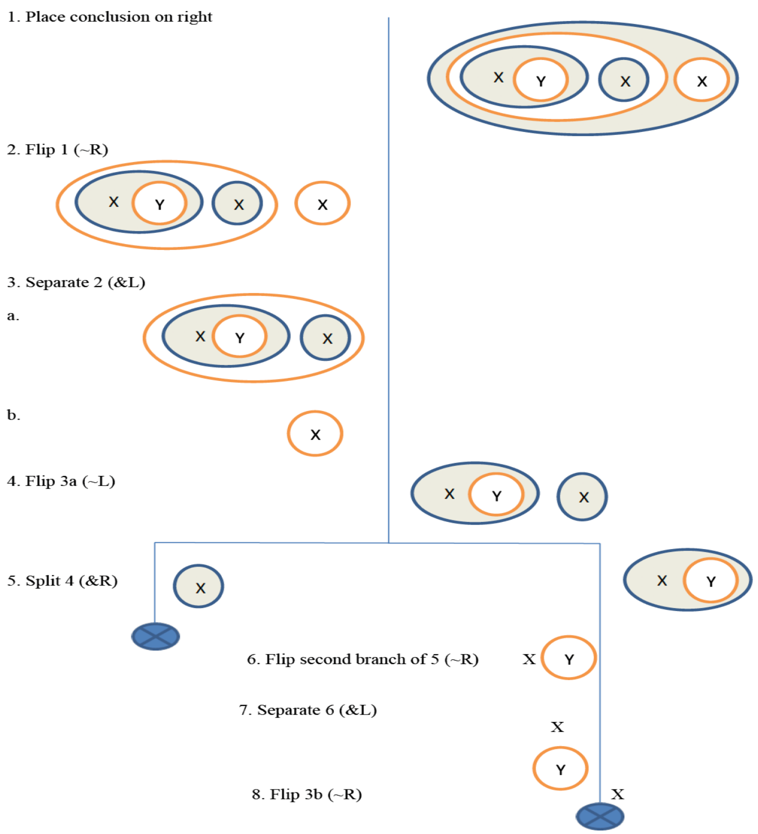


**Fig. 3.** Tree proof of Peirce's Law in graphical notation (read endoporeutically)

To be sure, this proof is slightly more involved than Peirce's reasoning. That is because these particular rules force us to understand a conditional in the graphical notation "endoporeutically" - as a negation of a conjunction (of antecedent and negated consequent). By contrast, Peirce's own justification for the law - as well as the corresponding bilateral tree proof in standard symbolic notation - directly applies the decomposition rule for conditionals. One can easily fix this shortcoming by adopting a multiple-reading understanding of the graphs ([12], Chap. 3), and then supplementing the set of rules with appropriate bilateral rules for the other sentential operations that may then be read off of them. Figures 4 and 5 gives us two such alternative bilateral tree proofs of the law, the first of which recapitulates Peirce's justification for the law just as directly as a bilateral tree proof in standard symbolic form. The second is noteworthy in that it treats the graphical representation of the law as a disjunction rather than a conditional, and proceeds by applying the relevant rule for a disjunction on the right. Since multiple readings may be extracted out of a single graph, they do not decompose in a unique fashion. As a result, such a system might yield bilateral tree proofs that are even *shorter* in the graphical notation than in the standard symbolic notation.[5]

One might have noticed that these rules for decomposing alpha graphs resemble the rules of the "endoporeutic" game Peirce envisions being played between a graphist and a grapheus, as they go about simplifying a given graph to determine whether or not it is true (against a set of antecedent constraints representing a given model or truth-value assignment) [3, 6, 11]. As Peirce shows, a given graph is true (according to a model) just in case the graphist (or proposer) has a winning strategy against the grapheus (or skeptic). Thus we can see Peirce here implementing in diagrams his long-standing idea that truth is that which would survive ultimate challenge. As in Peirce's endoporeutic, the rules of the proof system just described have us stripping off negations and handing graphs over to the other side. In effect, these bilateral tree diagrams illustrate the space of outcomes of a game strikingly reminiscent of Peirce's game, whereby a skeptic (the Grapheus) searches for a possible counterexample to a given argument form or sequent. In this game, players taking on the role of a skeptic have choices about which graph elements they wish to pursue in order to latch onto a particular counterexample. Those choices are then represented by branches in the bilateral proof diagram. If the Grapheus has a winning strategy, then a counterexample to that sequent form exists. The tree diagram effectively reduces a complex graph to elements inscribed on the recto (the left) and the verso (the right). It is as if one is viewing all the possibilities the game might play out from "sideways on."

About the same time he was developing his graphical system of logic, Peirce was also attempting to give a logical "proof" of the correctness of pragmatism that would "leave no reasonable doubt on the subject, and to be the one contribution of value that he has to make to philosophy." (CP 5.415) Peirce clearly thought the two projects to be interrelated. "I beg leave, Reader, as an Introduction to my defense of pragmatism, to bring before you a very simple system of diagrammatization of propositions which I term the System of Existential Graphs" (CP 4.534). For Peirce, a chief virtue of this system is that it allows the study of reasoning to escape psychologism and to become a normative science of diagrams in its own right.

> Diagrammatic reasoning is the only really fertile reasoning. If logicians would only embrace this method, we should no longer see attempts to base their science on the fragile foundations of metaphysics or a psychology not based on logical theory; and there would soon be such an advance in logic that every science would feel the benefit of it (CP 4.571).

---

[5] Consider, for instance, the respective proofs of conditional exchange and the DeMorgan's equivalencies.

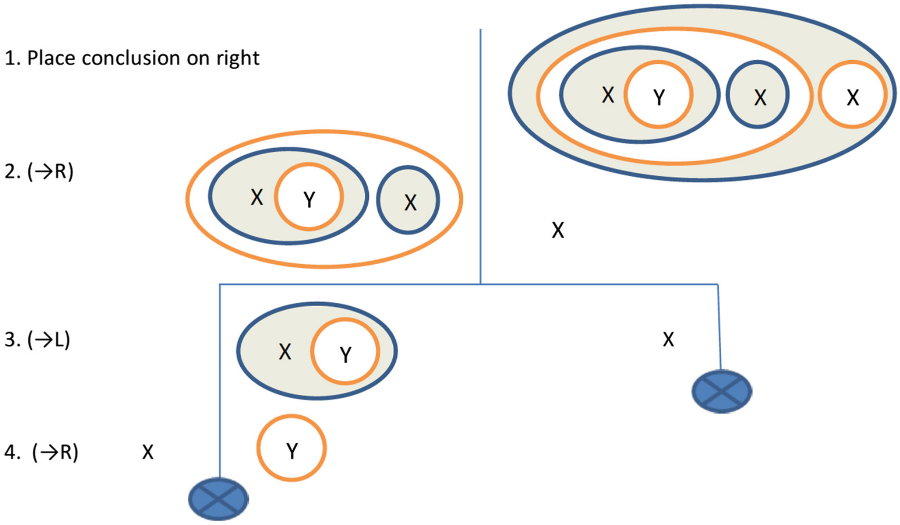


**Fig. 4.** Tree Proof of Peirce's Law (read as a conditional)

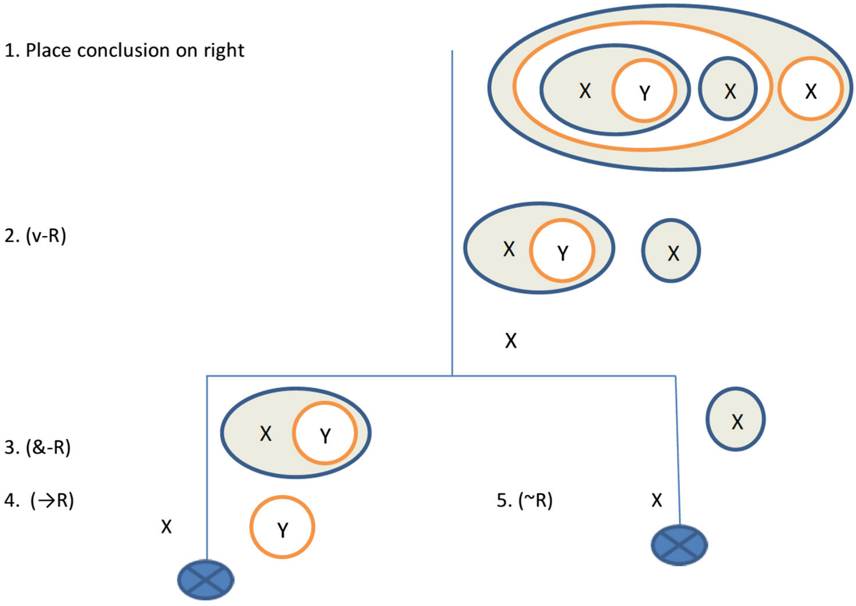


**Fig. 5.** Tree Proof of Peirce's Law (read as a disjunction)

Here it pays to look at the specific formulations of the pragmatic maxim for which he was actually trying to provide a proof. Consider, for instance, the following from the same *Monist* series of 1904–5 in which he introduces his existential graphs:

> Endeavoring, as a man of that type naturally would, to formulate what he so approved, he framed the theory that a conception, that is, the rational purport of a word or other expression, lies exclusively in its conceivable bearing on the conduct of life; so that, since obviously nothing that might not result from experiment can have direct bearing on the conduct, if one can define accurately all the conceivable experimental phenomena which the affirmation or denial of a concept could imply, one will have therein a complete definition of the concept, and there is absolutely nothing more in it (CP 5.412).

This articulation of the maxim is strikingly different from earlier expressions, in that it directs us to look at the consequences of affirming (and denying) claims, rather than emphasizing the (largely sensible) consequences of a claim's being true.[6] Another example from a 1904 review of Nichols' *Cosmology* is even more striking, in that it explicitly minimizes the role that sensation plays in the pragmatic maxim:

> The word pragmatism was invented to express a certain maxim of logic, which, as was shown at its first enouncement, involves a whole system of philosophy. The maxim is intended to furnish a method for the analysis of concepts. A concept is something having the mode of being of a general type which is, or may be made, the rational part of the purport of a word. A more precise or fuller definition cannot here be attempted. The method prescribed in the maxim is to trace out in the imagination the conceivable practical consequences,–that is, the consequences for deliberate, self-controlled conduct,–of the affirmation or denial of the concept; and the assertion of the maxim is that herein lies the whole of the purport of the word, the entire concept. The sedulous exclusion from this statement of all reference to sensation is specially to be remarked (CP 8.191).

Such an exclusion is clearly not observed in Peirce's original articulations and illustrations of the pragmatic maxim (the ones most widely repeated as canonical expressions of pragmatism), which have far more psychological elements. While Peirce doesn't comment on the reasons he makes these specific adjustments to the maxim's formulation, it stands to reason that these differences are at least partly driven by his desire to accommodate it to a logical proof – not an empirical demonstration. In moving from the sensible effects of a claim's being true to the consequences that redound to a speaker by affirming (and denying) a claim, Peirce is evidently moving away from an understanding of pragmatist semantics that aligns it with some form of verificationism. Such distancing from crude verificationism allows the maxim to avoid familiar self-application problems that plague expressions of verificationist principles of meaning. In so doing, Peirce allows for the maxim to be vindicated, not by its having specific empirical content, but rather by having *practical* application.

If that is so, however, then there is another striking feature of these 1904–5 formulations of the pragmatic maxim that has largely gone unnoticed by commentators [7, 11]. That would be their bilaterality; they direct us to look not only at the consequences of *affirming* claims, but also to the consequences of *denying* them. It is quite possible that these bilateral formulations of the pragmatic maxim are meant to map onto features of his system of logical notation, which we have seen to be similarly bilateral – as

---

[6] By contrast, compare this formulation with the one Peirce gives in his *Baldwin's Encyclopedia* entry on pragmatism (1902): "In order to ascertain the meaning of an intellectual conception one should consider what practical consequences might conceivably result by necessity from the truth of that conception; and the sum of these consequences will constitute the entire meaning of the conception" (CP 5.9).

comprised of flip sides of assertion and denial. That is, in speaking of the consequences of affirmation and denial, Peirce is thinking about claims and concepts as they appear on his existential graphs. The success of his graphs, then, might then provide some sort of vindication for his pragmatic maxim, so formulated. But as we have seen, this very bilaterality is incorporated much more directly in bilateral *tree* systems of proof than in his own system of permissions. For bilateral tree proofs explicitly show how tracing the consequences of both affirming premises and denying conclusions factor into determining whether and how entailment relations exist between claims. As a result, bilateral tree proofs literally *embody* these bilateral formulations of the pragmatic maxim, and that in turn could pave the way for a much more direct *pragmatic* justification for these expressions of pragmatism.

Unfortunately the promised proof of pragmatism, which Peirce told us (CP 4.572) to expect in a fourth article of the 1905 Monist series, never appears. Any reconstruction of how that demonstration is (or was) supposed to go has been left as a largely frustrating exercise for his commentators [7]. In any event, by 1907 Peirce seems to have given up his attempts to prove the maxim by means of his graphs, opting instead for a strategy that leans more heavily upon his general theory of signs. Pietarinen [11] suggests that this is likely due to the graphs' inability to represent certain features of modal reasoning, which Peirce stresses would need to be captured in order to accommodate appropriately subjunctive readings of the pragmatic maxim. However, that only points to yet another advantage of tree systems, which are readily extended to modal reasoning [5].

For these reasons, I close by surmising that had Peirce only been aware of tree methods of proof, he might not have been so quick to abandon his efforts to demonstrate the truth of his pragmaticist formulations of the pragmatic maxim, by way of orienting or incorporating them in a system of logical diagrams. Herein lies something of a sad ironic twist. While the development of Tableau systems of proof are mostly associated with E. W. Beth and Jaakko Hintikka in the 1950's, and then popularized by Jeffreys and Smullyan in the 1960's, they actually go back to one of Peirce's contemporaries [2]. In 1894, Lewis Carroll devised a tree method to solve extended, multiliteral sorites puzzles. As Carroll wrote in his diary on July 16, 1894:

> Today has proved to be an epoch in my Logical work. It occurred to me to try a complex Sorites by the method I have been using for ascertaining what cells, if any, survive for possible occupation when certain nullifies are given. I took one of 40 premises, with 'pairs within pairs,' & many bars, & worked it like a genealogy, each term proving all its descendents. It came out beautifully, & much shorter than the method I have used hitherto - I think of calling it the 'Genealogical Method' ([4], p. 279).

Eventually, Carroll settled on labeling his discovery "the method of trees," describing it as follows:

> In the Method of Trees this process is reversed. Its essential feature is that it involves a Reductio ad Absurdum. That is, we begin by assuming, argumenti gratia, that the aggregate of the Retinends (which we wish to prove to be a Nullity) is an Entity: from this assumption we deduce a certain result: this result we show to be absurd: and hence we infer that our original assumption was false, i.e., that the aggregate of the Retinends is a Nullity ([4], p. 280).

That is, Carroll's method has one hypothesize a possible object with a set of attributes and non-attributes determined by endpoints generated from the premises of a given sorites. The method then systematically applies the various premises within the sorites to determine what further attributes and non-attributes such an object would have to have. Occasionally, such determinations require us to branch paths and consider various alternate combinations of attributes and non-attributes. I take such path branching to be the key diagrammatic innovation behind all tree systems of proof. Like contemporary tree systems, the ultimate aim is to close off branches, either by deducing that the hypothesized object would have to possess both an attribute and its negation or by showing that such a combination is explicitly ruled out by a premise. Accordingly, Carroll's system has two negations built into it, one (signified by a ') representing the non-possession of an attribute, the other (signified by a 0) representing the non-existence of a certain class of object. In the end, the full closure of an entire tree signifies that we may conclude the non-existence of the originally hypothesized object.

The story of Carroll's "method of trees" is a tale worthy of its own telling. Originally slated to be included in Volume II of his *Symbolic Logic*, Carroll sent a description of the method to John Cook Wilson in 1896. Wilson had not returned the copy to Carroll by the time of his death in 1898. So while most of the material in Carroll's Oxford office was subsequently tossed out with the rubbish, fortunately a relatively complete description of Carroll's method of trees survived. However, it largely remained lost until William Bartley III brought the method back to light in Book XII of his attempted reconstruction of the second volume of Carroll's *Symbolic Logic* [4]. In the meantime, Hintikka, Beth and their followers had independently developed tree methods with branching features similar to those that Carroll had devised earlier.

Carroll was evidently aware of the attempts by Peirce and his students at Johns Hopkins to develop visual or diagrammatic methods of demonstrating argumentative validity. Carroll's personal library included a copy of their 1883 *Studies in Logic* [9], and it has been suggested that the inference engine driving Carroll's method of trees is inspired by Christine Ladd-Franklin's inconsistent triad (or antilogism) [1]. That Carroll was aware of Ladd-Franklin is evident from the several exercises he "borrows" from her.[7] However, Carroll's logical sensibilities were dramatically different from Peirce's. Centered as it was around categories and not propositions, Carroll's logic would have struck Peirce as still quaintly wed to the Aristotelian tradition. He would have been unimpressed by Carroll's use of multiple flavors of negation, and even less impressed by his adherence to the idea that universal claims have existential import. So it is likely that the recognition and admiration was not mutual, especially when one considers Peirce's disparaging remarks about "overcultivated Oxford Dons – I hope their day is over" (CP 5.520). That is a shame, for those commitments about logic that divide Peirce and Carroll can easily be pried apart from the type of logical diagram that Carroll was developing. Had Peirce only been (more?) appreciative of, or even familiar

---

[7] Amirouche Moktefi has pointed out to me (in correspondence) that this raises the question of whether Peirce himself appreciated the originality of Franklin-Ladd's antilogism. As noted above, Peirce's system of permissions does not take advantage of potential incompatibilities in the way that tree systems (including Carroll's) do.

with, the work of the estimable Charles Dodgson, then I imagine that the method of trees would not have been lost until their rediscovery in the mid 1900s, and logic might have taken a more visual, and less symbolic, turn.[8]

## References

1. Abeles, F.: Lewis Carroll's method of trees: its origins in studies in logic. Mod. Log. **1**, 25–35 (1990)
2. Anellis, I.: From semantic tableaux to smullyan trees: a history of the development of the falsifiability tree method. Mod. Log. **1**, 36–69 (1990)
3. Burch, R.: Game-theoretical semantics for peirce's existential graphs. Synthese **99**, 361–375 (1994)
4. Carroll, L.: Symbolic Logic, Part I and Part II. Bartley III., W.W. (ed.). Clarkson N Potter, New York (1977)
5. Garson, J.: Modal Logic for Philosophers. Cambridge University Press, Cambridge (2006)
6. Hilpinen, R.: On C. S. Peirce's theory of the proposition: Peirce as a precursor of game-theoretical semantics. Monist **65**, 182–188 (1982)
7. Hookway, C.: The Pragmatic Maxim. Oxford University Press, Oxford (2012)
8. Misak, C.: Cambridge Pragmatism. Oxford University Press, Oxford (2017)
9. Moktefi, A.: Are other people's books difficult to read? The logic books in Lewis Carroll's private library. Acta Baltica Historiae et Philosophiae Scientarum **5**(1), 28–49 (2017)
10. Peirce, C.S.: The Collected Papers of Charles S. Peirce, 8 volumes, Hartshorne, C., Weiss, P., Burns, A.W. (eds.). Harvard University Press, Cambridge (1931–1966). (References to Peirce's papers are designated by CP followed by volume and paragraph number)
11. Pietarinen, A.-V.: Moving pictures of thought II: graphs, games, and pragmaticism's proofs. Semiotica **186**, 315–331 (2011)
12. Shin, S.-J.: The Iconic Logic of Peirce's Graphs. MIT Press, Cambridge (2002)
13. Sieg, W.: Logic and Proofs. Course available through the Open Learning Initiative. http://oli.cmu.edu/courses/all-oli-courses/logic-proofs-course-details/
14. Sowa, J.: Peirce's tutorial on existential graphs. Semiotica **186**, 345–394 (2011)

---

[8] While it is rather hard to imagine how Peirce could have been exposed to Carroll's later work on logical diagrams, there nevertheless was at most only two degrees of separation between them. Not only did Peirce maintain a substantial correspondence with Victoria Welby from 1903–11, Welby (who championed Peirce's thought in the UK) is also known to have corresponded with Cook Wilson (see [8], p. 83). That then might be one place to look for any hint of a flow of information back from Carroll to Peirce.

# A Weakening of Alpha Graphs: Quasi-Boolean Algebras

Minghui Ma[1] and Ahti-Veikko Pietarinen[2,3](✉)

[1] Institute of Logic and Cognition, Sun Yat-sen University, Guangzhou, China
mamh6@mail.sysu.edu.cn
[2] Tallinn University of Technology, Tallinn, Estonia
ahti-veikko.pietarinen@ttu.ee
[3] Nazarbayev University, Astana, Kazakhstan

**Abstract.** Peirce introduced the Alpha part of the logic of Existential Graphs (EGs) as a diagrammatic syntax and graphical system corresponding to classical propositional logic. The logic of quasi-Boolean algebras (De Morgan algebras) is a weakening of classical propositional logic. We develop a graphical system of weak Alpha graphs for quasi-Boolean algebras, and show its soundness and completeness with respect to this algebra. Weak logical graphs arise with only minor modifications to the transformation rules of the original theory of EGs. Implications of these modifications to the meaning of the sheet of assertion are then also examined.

**Keywords:** Existential Graphs · Weak Alpha graphs
Quasi-Boolean algebra · Sheet of Assertion

## 1 Introduction

Alpha graphs are the first part of the theory of Existential Graphs (EGs) which Charles Peirce developed in 1896. Named as Alpha and systematically investigated in his 1903 Lowell Lectures (see e.g. R 450, R 454, R 455–455),[1] most of Peirce's ideas have remained unpublished despite the fact that he studied his graphical method extensively in his later manuscripts. The second part of EGs is the first-order quantificational logic (Beta graphs, [22]) and the third the modal and higher-order logic (Gamma graphs). Peirce also mentioned, but did not develop, the fourth, Delta part. As a logical syntax, EGs are two-dimensional graphs, and in contrast to linear notations are sometimes considered as having

---

M. Ma—The work supported by the Project Supported by Guangdong Province Higher Vocational Colleges & Schools Pearl River Scholar Funded Scheme (2017).
A.-V. Pietarinen—The work supported by the Estonian Research Council Personal Research Grant PUT 1305 (*Abduction in the Age of Fundamental Uncertainty*) and Nazarbayev University Social Policy Grant 2018–2019.

[1] The reference R is to Peirce's manuscripts by the Robin number [19].

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 549–564, 2018.
https://doi.org/10.1007/978-3-319-91376-6_50

the nature of diagrams [21]. The theory of Alpha graphs as classical propositional logic has been studied in [1,20,24–27], among others.

Alpha graphs emerged when Peirce was developing the graphical method of logic arising from his work on the algebra of the copula (R 411, 417, 482, [2]). His work in mathematical logic began with the improvement of Boole's algebra in 1867 [15]. Boole's operations of negation and disjunction were modified to become operations in contemporary Boolean algebras. Later, in his 1880 work on the algebra of logic [16], Peirce proposed the first sound and complete system for Boolean algebras in the history of mathematical logic.[2] This work was soon followed by his paper on the algebra of logic published in 1885 [17], taken as a contribution to the philosophy of notation. Peirce continued his work on Boolean algebras in the 1890s, culminating in his invention of EGs in 1896. The purpose of this method was to divide necessary, mathematical reasoning into its ultimate elementary logical steps, and to analyze those steps and their composition by the special rules of inference of the graphical system.

We mention this algebraic preface to his graphical or diagrammatic phase of logical investigations for the following reason. Peirce recognized the power of the algebraic way of thinking in guiding the progress of logical investigations. This is communicated well in Peirce's response to the letter he had received from his former student, Christine Ladd–Franklin, who had written to Peirce in order to ask his opinion of what the study of graphs would add to the study of the algebraic approach that they already had at their disposal. Peirce's reply is worth noting: "You ask whether Logical Graphs have any bearings on Non-Relative logic. Not *much, except* in one highly important particular, that they supply an entirely new system of fundamental assumptions to logical algebra" (November 9, 1900, Peirce's own emphasis).

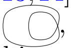
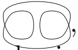
As we will observe in this paper, Peirce was right in the recognition of the value of logical graphs in the following sense. It is possible not only to provide algebraic rules and algebraic semantics to the theory of Alpha graphs (on this, see [11]), but also to develop other graphical logics besides classical Alpha graphs or extensions to the Beta and the Gamma parts, based on the underlying algebraic motivations of the theory (see e.g. [13,22]). In the present paper, we show how to weaken the Alpha graphs from classical bivalent propositional logic to obtain a four-valued logic. One of the algebras of that logic is known as a quasi-Boolean or De Morgan algebra [5,7].

Peirce's theory of EGs is not only a theory of logical syntax and semantics but also a proof theory for many logics. In a sense his "entirely new system of fundamental assumptions" that logical graphs add to logical algebra is uncovered in the different way inferential relations can be conceived of in the two-dimensional diagrammatic syntax of logical graphs. A set of transformation rules of graphs defines a proof system for a graphical logic. Now Brünnler [6] has developed a *deep inference* system for classical propositional logic in which a rule can be applied at any position in a deep structure. Ma and Pietarinen [11] reconstructed

---

[2] He presented the 1880 system as an instrument for the analysis of the logical consequence relation, and thus came close to a sequent-style calculus in that presentation.

a graphical system of Alpha graphs for classical propositional logic which has the nature of deep inference. The graphical and diagrammatic syntax of graphs allows to apply transformation rules at any position of a graph. This is exactly the deep inference methodology, where inference rules apply not only on outermost connectives of formulas but also arbitrary deep inside them.

For the purposes of the present study, the rules of transformation are modified to achieve desired effects that can breed novel graphical logics. Our proposed modification here results in a weakening of the classical logic—or, alternatively, a strengthening of an implication-free intuitionistic logic. The intuitionistic logic of Alpha graphs has been developed in [4,13,14]. There the graphical language has two new primitives: the *scroll*, notation , which corresponds to intuitionistic implication, and the *double-scroll* (and its generalisations), notation , which corresponds to intuitionistic disjunction. In the present paper, we begin with the language of the Alpha graphs which uses not these scroll notations but simple closed curves or ovals, notation , together with the nesting of ovals around any graph.

The resulting expressions of weak Alpha graphs thus look just like what the syntax of Peirce's original theory of Alpha graphs would produce. The difference is that in our weakening of the Alpha there is no sign of implication: the nested structures of ovals can only be interpreted either as conjunctions or as disjunctions. Also, ovals do not have the meaning they have in original Alpha graphs, namely that of the classical negation: although the double-oval rule  $= P$ is valid in a quasi-Boolean algebra, it does not imply the presence of a classically interpreted rule of double negation. The meaning of the double-oval rule is that of an algebraic operation of *involution*. The oval is a dual endomorphism operator. These weak Alpha graphs thus preserve a natural duality: there is the Sheet of Assertion SA on which the graphs are scribed, and the blank SA is the top element of the algebra. An oval drawn on the blank SA is the dual of the top element, namely the bottom. Unlike in intuitionistic graphs, in which $\neg(P \wedge Q) \to \neg P \vee \neg Q$ is not valid, both of the De Morgan rules come out as valid in the system of these weakened Alpha graphs. Hence the name 'De Morgan' is sometimes used instead of 'quasi-Boolean' algebras.

The next section presents quasi-Boolean algebras in a logical perspective. Section 3 presents the new language and the system for weak Alpha graphs. Section 4 proves the soundness and completeness of the system with respect to quasi-Boolean algebras. Section 5 puts the proposal into a perspective and proposes an application to machine-based reasoning methods. Section 6 concludes.

## 2 The Logic of Quasi-Boolean Algebras

In this section, we present the essentials of quasi-Boolean algebras in the logical perspective. Throughout, we assume familiarity with the basics of Peirce's theory of EG, especially those of its Alpha part that corresponds to the theory of classical propositional logic.

**Definition 1.** *A* bounded distributive lattice *is an algebra* $\mathbb{A} = (\mathfrak{A}, \wedge, \vee, 0, 1)$ *satisfying the following conditions for all* $a, b, c \in \mathfrak{A}$*:*

– *Lattice laws:*

$$\begin{array}{ccr} a \wedge b = b \wedge a & a \vee b = b \vee a & \text{(Commutativity)} \\ a \wedge (b \wedge c) = (a \wedge b) \wedge c & a \vee (b \vee c) = (a \vee b) \vee c & \text{(Associativity)} \\ a \wedge a = a & a \vee a = a & \text{(Idempotency)} \\ a \wedge (a \vee b) = a & a \vee (a \wedge b) = a & \text{(Absorption)} \end{array}$$

– *Distributivity:* $a \wedge (b \vee c) \leq (a \wedge b) \vee (a \wedge c)$.
– *Boundedness:* $a \leq 1$ *and* $0 \leq a$.

*The order* $\leq$ *is defined by:* $a \leq b$ *if and only if* $a = a \wedge b$*, or equivalently* $a \vee b = b$.

*A* quasi-Boolean algebra *(*qBa *for short) is an algebra* $\mathbb{A} = (\mathfrak{A}, \wedge, \vee, \sim, 0, 1)$ *where* $(\mathfrak{A}, \wedge, \vee, 0, 1)$ *is a bounded distributive lattice and* $\sim$ *is a unary operation on* $\mathfrak{A}$ *satisfying the following condition for all* $a, b \in \mathfrak{A}$*:*

– De Morgan Laws*:* $\sim(a \vee b) = \sim a \wedge \sim b$ *(the first De Morgan Law) and* $\sim(a \wedge b) = \sim a \vee \sim b$ *(the second De Morgan Law).*[3]
– Involution*:* $\sim\sim a = a$.
– $\sim 0 = 1$ *and* $\sim 1 = 0$.

*Let* qBa *be the variety of all quasi-Boolean algebras.*

Let BA be the variety of Boolean algebras. It is clear that a quasi-Boolean algebra $\mathbb{A} = (\mathfrak{A}, \wedge, \vee, \sim, 0, 1)$ is a Boolean algebra if and only if $a \vee \sim a = 1$ for all $a \in \mathfrak{A}$. Hence BA is a subvariety of qBa.

**Definition 2.** *Let* $\mathcal{V}$ *be a denumerable set of variables. The set of all terms* $\mathcal{T}$ *is defined inductively by the following rule:*

$$\mathcal{T} \ni \phi ::= p \mid \bot \mid \top \mid (\phi_1 \wedge \phi_2) \mid (\phi_1 \vee \phi_2) \mid \sim\phi$$

*where* $p \in \mathcal{V}$*. A* sequent *is an expression of the form* $\Gamma \Rightarrow \phi$ *where* $\Gamma$ *is a non-empty finite multi-set of terms and* $\phi$ *is a term. For any nonempty finite multi-set of terms* $\Gamma = \{\phi_1, \ldots, \phi_n\}$*, let* $f(\Gamma) = \phi_1 \wedge \ldots \wedge \phi_n$.

**Definition 3.** *The Gentzen sequent calculus* GqBa *for quasi-Boolean algebras consists of the following axioms and rules:*

*(1) Axioms:*

$$(\mathrm{Id}_1)\ p, \Gamma \Rightarrow p \quad (\mathrm{Id}_2)\ \sim p, \Sigma \Rightarrow \sim p$$

$$(\bot\Rightarrow)\ \bot, \Gamma \Rightarrow \phi \quad (\Rightarrow\sim\bot)\ \Gamma \Rightarrow \sim\bot$$

---

[3] The De Morgan law $\neg(a \wedge b) \leq \neg a \vee \neg b$ with intuitionistic negation $\neg$ does not hold in Heyting algebras while its converse $\neg a \vee \neg b \leq \neg(a \wedge b)$ does. However, both De Morgan laws hold in quasi-Boolean algebras.

*(2) Logical rules:*

$$\frac{\phi, \psi, \Gamma \Rightarrow \chi}{\phi \wedge \psi, \Gamma \Rightarrow \chi}(\wedge{\Rightarrow}) \qquad \frac{\Gamma \Rightarrow \phi \quad \Gamma \Rightarrow \psi}{\Gamma \Rightarrow \phi \wedge \psi}({\Rightarrow}\wedge)$$

$$\frac{\phi, \Gamma \Rightarrow \chi \quad \psi, \Gamma \Rightarrow \chi}{\phi \vee \psi, \Gamma \Rightarrow \chi}(\vee{\Rightarrow}) \qquad \frac{\Gamma \Rightarrow \phi_i}{\Gamma \Rightarrow \phi_1 \vee \phi_2}({\Rightarrow}\vee)(i = 1, 2)$$

$$\frac{\sim\phi, \Gamma \Rightarrow \chi \quad \sim\psi, \Gamma \Rightarrow \chi}{\sim(\phi \wedge \psi), \Gamma \Rightarrow \chi}({\sim}\wedge{\Rightarrow}) \qquad \frac{\Gamma \Rightarrow \sim\phi_i}{\Gamma \Rightarrow \sim(\phi_1 \wedge \phi_2)}({\Rightarrow}{\sim}\wedge)(i = 1, 2)$$

$$\frac{\sim\phi, \sim\psi, \Gamma \Rightarrow \chi}{\sim(\phi \vee \psi), \Gamma \Rightarrow \chi}(\sim \vee \Rightarrow) \qquad \frac{\Gamma \Rightarrow \sim\phi \quad \Gamma \Rightarrow \sim\psi}{\Gamma \Rightarrow \sim(\phi \vee \psi)}({\Rightarrow}{\sim}\vee)$$

$$\frac{\phi, \Gamma \Rightarrow \chi}{\sim\sim\phi, \Gamma \Rightarrow \chi}({\sim}{\sim}{\Rightarrow}) \qquad \frac{\Gamma \Rightarrow \phi}{\Gamma \Rightarrow \sim\sim\phi}({\Rightarrow}{\sim}{\sim})$$

*A* derivation *in* GqBa *is a tree-like structure $\mathcal{D}$ where each node is either an axiom or derived by a rule from the child node(s). The* height *of a derivation $\mathcal{D}$ is the length of the longest branch of $\mathcal{D}$. The notation* GqBa $\vdash \Gamma \Rightarrow \psi$ *stands for that $\Gamma \Rightarrow \psi$ is derivable in* GqBa*. A sequent rule is a fraction*

$$\frac{\Gamma_1 \Rightarrow \psi_1 \ldots \Gamma_n \Rightarrow \psi_n}{\Gamma_0 \Rightarrow \psi_0}(R)$$

*where $\Gamma_1 \Rightarrow \psi_1, \ldots, \Gamma_n \Rightarrow \psi_n$ are called* premisses *of $(R)$, and $\Gamma_0 \Rightarrow \psi_0$ is called the* conclusion *of $(R)$. A sequent rule $(R)$ is* admissible *in* GqBa*, if the conclusion of $(R)$ is derivable in* GqBa *whenever the premisses of $(R)$ are derivable in* GqBa*.*

*Example 1.* The distributivity law $\phi \wedge (\psi \vee \chi) \Rightarrow (\phi \wedge \psi) \vee (\phi \vee \chi)$ is derived in GqBa as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\phi, \psi \Rightarrow \phi \quad \phi, \psi \Rightarrow \psi}{\phi, \psi \Rightarrow \phi \wedge \psi}({\Rightarrow}\wedge)}{\phi, \psi \Rightarrow (\phi \wedge \psi) \vee (\phi \wedge \chi)}({\Rightarrow}\vee) \qquad \dfrac{\dfrac{\phi, \chi \Rightarrow \phi \quad \phi, \chi \Rightarrow \chi}{\phi, \chi \Rightarrow \phi \wedge \chi}({\Rightarrow}\wedge)}{\phi, \chi \Rightarrow (\phi \wedge \psi) \vee (\phi \wedge \chi)}({\Rightarrow}\vee)}{\phi, \psi \vee \chi \Rightarrow (\phi \wedge \psi) \vee (\phi \wedge \chi)}(\vee{\Rightarrow})}{\phi \wedge (\psi \vee \chi) \Rightarrow (\phi \wedge \psi) \vee (\phi \wedge \chi)}(\wedge{\Rightarrow})$$

Other distributive laws can be derived in GqBa similarly.

**Theorem 1.** *The following cut rule is admissible in* GqBa*:*

$$\frac{\Gamma \Rightarrow \phi \quad \phi, \Delta \Rightarrow \psi}{\Gamma, \Delta \Rightarrow \psi}(cut)$$

*Proof.* The proof proceeds by simultaneous induction on the heights of derivations of the premisses of (*cut*) and the complexity of the cut term $\phi$. The details can be found in [9]. □

An *assignment* in a qBa $\mathbb{A} = (\mathfrak{A}, \wedge, \vee, \sim, 0, 1)$ is a function $\theta\colon \mathcal{V} \to \mathfrak{A}$. An assignment $\theta$ in $\mathbb{A}$ can be extended homomorphically to the term algebra on $\mathcal{T}$. For any term $\phi \in \mathcal{T}$, let $\theta(\phi)$ be the value of $\phi$ in $\mathfrak{A}$ under the assignment $\theta$ in $\mathbb{A}$. A sequent $\Gamma \Rightarrow \psi$ is *valid* in $\mathbb{A}$, notation $\mathbb{A} \models \Gamma \Rightarrow \psi$, if $\theta(f(\Gamma)) \leq \theta(\psi)$ for any assignment $\theta$ in $\mathbb{A}$. We say that a sequent $\Gamma \Rightarrow \psi$ is *valid* in qBa, notation qBa $\models \Gamma \Rightarrow \psi$, if $\mathbb{A} \models \Gamma \Rightarrow \psi$ for all $\mathbb{A} \in$ qBa.

**Theorem 2.** GqBa $\vdash \Gamma \Rightarrow \psi$ *if and only if* qBa $\models \Gamma \Rightarrow \psi$*.*

*Proof.* The soundness part is shown easily by induction on the height of a derivation of $\Gamma \Rightarrow \psi$. The completeness part is shown easily by standard Lindenbaum–Tarski construction. The details can be found in [9, Theorem 4.11]. □

## 3 A Graphical System for qBas

Peirce introduced the Alpha graphs and a graphical system for deductive inferences for Boolean algebras in a number of manuscripts (R 480–482, cf. [11]). In November 1896 he presented his innovation at the National Academy of Sciences meeting held in Columbia College, New York City (R 488).

The logic of qBas is a weakening of the classical propositional logic. In order to develop a graphical system for the logic of qBas, we shall use the geometrical (or the diagrammatic) syntax of *Alpha graphs* which are graphical expressions scribed on the *Sheet of Assertion* (SA). The SA is a two-dimensional, open-compact space uniform in all directions. What does the SA express? Here we take the SA to mean the top element $\top$ in a qBa, without any further significance attached to it.

A variable in $\mathcal{V}$ or the SA is called an *atomic graph*. For the construction of an Alpha graph, we may write an atomic graph on the SA or start from the SA. An Alpha graph is constructed from atomic graphs using the operations of the *oval* and *juxtaposition*. The operation of the *oval* is to draw a simple closed curve ◯ around a graph such that it encloses another Alpha graph $G$ to form a new Alpha graph $\ovalbox{G}$ . In particular, we have the Alpha graph ◯ which is obtained from the SA by an application of the operation of drawing an oval on the SA. The main function of the oval is to partition the SA into areas. An *area* is a continuous space in a graph. For example, the Alpha graph

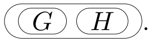
$$1\ \ovalbox{2\ \ovalbox{3\ \ovalbox{4\quad}}}$$

contains four areas (annotated here by numbers for clarity), from the outside-in direction. Juxtaposition is the operation which scribes two graphs $G$ and $H$ in the same area. Ovals are either nested or juxtaposed and cannot produce overlapping instances of Alpha graphs. Now we give the formation rules of Alpha graphs.

**Definition 4.** *The set of all* Alpha graphs $\mathfrak{G}_\alpha$ *is constructed from variables or from the* SA *by using the following inductive rule:*

$$\mathfrak{G}_\alpha \ni G ::= p \mid \mathsf{SA} \mid G\,H \mid \ovalbox{G}$$

*where* $p \in \mathcal{V}$*. The Alpha graph* $G\,H$ *is obtained from* $G$ *and* $H$ *by the operation of juxtaposition. The graph*  *is obtained from* $G$ *by the operation of encircling by the oval. Let us define* $G \circledvee H$ *as the graph:*

.

*The graph* $G \circledvee H$ *is called the* graphical disjunction *of* $G$ *and* $H$*. A* graphical consequence *is of the form* $G \vdash H$ *where* $G$ *and* $H$ *are Alpha graphs.*

Every graph $G$ has a parsing tree $T_\alpha(G)$ which can be defined inductively (see [11, Definition 2.3]). A *partial graph* (a sub-graph) of $G$ is the graph of a node of the parsing tree $T_\alpha(G)$. An *entire graph* of $G$ is the graph labelled at the *root* of the $T_\alpha(G)$ and it is the graph on the $\mathsf{SA}$ juxtaposed with nothing but the blank.

For any graph $G$, each area in $G$ has a polarity. The polarity of an area is *positive* if it is enclosed within an even number of ovals, and it is *negative* if it is enclosed within an odd number of ovals. The $\mathsf{SA}$ occurs within a zero number of ovals and is thus positive. An area can be viewed as a continuum of points, or positions, and at each position in an area a graph can be scribed. A graphical *context* is a graph $G\{\ \}$ with a single *slot* $\{\ \}$, the empty context or an area in $G$. Let $G\{H\}$ be the graph obtained from $G\{\ \}$ by scribing $H$ on the slot area.

The slot in a graphical context is either positive or negative. The notations $G^+\{\ \}$ and $G^-\{\ \}$ stand for that the slot is a positive and negative area, respectively. Let $G^+\{J\}$ and $G^-\{J\}$ be graphs in which $J$ is in a positive area and in a negative area, respectively.

**Definition 5.** *A graphical rule is a fraction of the form*

$$\frac{G}{H}(R)$$

*where* $G$ *is called the* premiss *and* $H$ *is called the* conclusion *of* $(R)$*.*

The transformation rules of the graphical system for quasi-Boolean algebras are formulated in terms of deep structures in an Alpha graph. It follows that the graphical system is indeed a deep inference system in which a rule can be applied at an appropriate position in any graph. This gives graphical systems advantage over other, non-graphical methods, as the deep-inference nature of inference rules is a natural property of graph transformations. The advantage comes in especially handy when devising proof systems for various non-classical logics.

The primary rules can be used to identify some graphs. The following rules of commutativity and associativity come to mind:

$$\text{(CM) } G\{H_1H_2\} = G\{H_2H_1\}; \quad \text{(AS) } G\{H_1(H_2H_3)\} = G\{(H_1H_2)H_3\}.$$

The commutativity (CM) says that positions of $H_1$ and $H_2$ in $H_1H_2$ are immaterial. The associativity (AS) says that the order of forming graphs as indicated by parentheses is immaterial. Notice that (AS) can express order only in such linear notation that has parentheses among its alphabet. The complete commutativity of juxtaposition in graphs $G$ and $H$ can nevertheless be shown as the identity of the following graphs:

This shows that each area in a graph is symmetric in all directions. Associativity follows from this commutativity.[4] Also, a formulation of the rule of conjunction becomes unnecessary, since the continuity of the SA guarantees that the premisses that rest on the same area of the sheet are juxtaposed with each other.

**Definition 6.** *The system $\mathfrak{S}_{qBa}$ consists of the following graphical rules:*

- Deletion rule*:*

$$\frac{G^{+}\{H\}}{G^{+}\{\mathsf{SA}\}}\,(\mathrm{Del})$$

  *Every positive partial graph $H$ in a graph $G$ can be deleted, leaving the blank* SA *in its place.*
- Insertion rule*:*

$$\frac{G^{-}\{H\}}{G^{-}\{JH\}}\,(\mathrm{Ins})$$

  *Any graph can be inserted into a negative position in a graph $G$.*
- Positive Iteration/Positive Deiteration rules*:*

$$\frac{K\{GH^{+}\{J\}\}}{K\{GH^{+}\{GJ\}\}}\,(\mathrm{Pit}) \qquad \frac{K\{GH^{+}\{GJ\}\}}{K\{GH^{+}\{J\}\}}\,(\mathrm{Pdeit})$$

  *In any graph $K\{GH^{+}\{J\}\}$, the partial graph $G$ can be iterated in the context $H^{+}\{\ \}$. Conversely, any graph that is the result of such iteration can be deiterated.*
- Involution rule*:*

$$\frac{G\{H\}}{G\{\boxed{\boxed{H}}\}}\,(\mathrm{Inv1}) \qquad \frac{G\{\boxed{\boxed{H}}\}}{G\{H\}}\,(\mathrm{Inv2})$$

  *A doubly occurring nested oval with nothing but a blank between the two ovals can be added to or removed from any partial graph.*

*A* derivation *from a graph $G$ to $H$ in $\mathfrak{S}_{qBa}$ is a finite sequence of graphs $K_0, \ldots, K_n$ such that $K_0 = G$ and $K_n = H$ and each $K_i$ is obtained from some $K_j$ $(j < i)$ by a rule in $\mathfrak{S}_{qBa}$. A graphical consequence $G \vdash H$ is* derivable *in $\mathfrak{S}_{qBa}$, notation $G \vdash_{\mathfrak{S}_{qBa}} H$, if there is a derivation of $H$ from $G$ in $\mathfrak{S}_{qBa}$.*

*A graph $G$ is* equivalent *to $H$ with respect to $\mathfrak{S}_{qBa}$, notation $G \equiv_{\mathfrak{S}_{qBa}} H$, if $G \vdash_{\mathsf{qBa}} H$ and $H \vdash_{\mathfrak{S}_{qBa}} G$.*

---

[4] Peirce knew well this point: "[I]f an operation is thoroughly commutative ..., it is necessarily associative; and associative property is a mere corollary from its commutative property" (R 482).

*Remark 1.* There are some important points of difference between $\mathfrak{S}_{qBa}$ and the Alpha graphs under Peirce's original conception (see the definition of the Alpha system given e.g. in [11]). Among them we list the following:

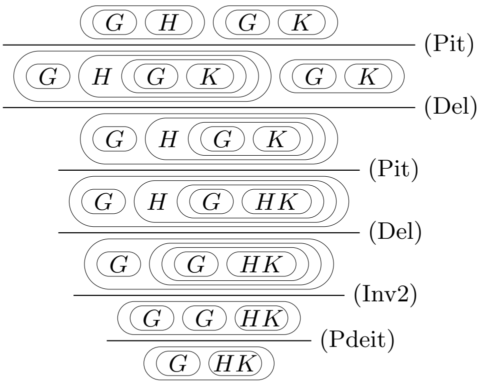
(1) The axiom of SA does not occur in $\mathfrak{S}_{qBa}$.
(2) No implication appears in the language of $\mathfrak{S}_{qBa}$. Although the definition of the well-formed graphs of $\mathfrak{S}_{qBa}$ is equivalent to the definition of the well-formed Alpha graphs, the systemic restrictions make it impossible to read graphs such as $(G\,(H))$ as representing implications. In terms of the language of propositional logic, this graph can be interpreted either as $\sim(G \wedge \sim H)$, $\sim(\sim H \wedge G)$, $(\sim G \vee H)$ or $(H \vee \sim G)$), and that is all.
(3) The iteration/deiteration rules familiar from the set of basic transformation rules of the Alpha system are restricted to *positive* iterations and deiterations as given by the rules of transformation (Pit) and (Pdeit). Only graphs on positive areas can be iterated, and only graphs iterated from positive areas can be deiterated. This restriction prevents the derivability of the Laws of the Excluded Middle as well as Non-Contradiction: SA $\nvdash_{\mathfrak{S}_{qBa}}$ $((G)\,((G)))$ and SA $\nvdash_{\mathfrak{S}_{qBa}}$ $(G\,(G))$. The proof by completeness is given at the end of the next section.
(4) The oval around graphs, $(G)$, does not represent a Boolean negation. Taking it as complementation would make $\mathfrak{S}_{qBa}$ a system for Boolean algebra, in which case we would call the oval the *cut*. But here, the oval is involution expressing duality. We do not cut what the oval encloses from the SA.
(5) The meaning of the Sheet of Assertion SA also differs from the classical Alpha. In $\mathfrak{S}_{qBa}$ it is a sheet, that is, it provides the geometrical, two-dimensional surface upon which graphs are projected. But in contrast to the proposals in the previous literature (including Peirce's own writings), here the sheet is no longer the Sheet of *Assertion*. That is, we will divest the SA from its significance of being about assertions. This conceptual change creates some potential for applications proposed in Sect. 5.

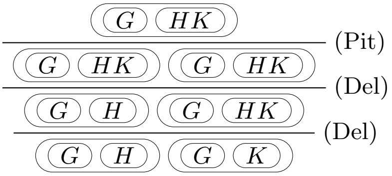
**Proposition 1.** *The following distributive laws are derivable in* $\mathfrak{S}_{qBa}$*:*

$$\text{(D1)}\ G\,((H)\,(K)) \vdash ((GH)\,(GK))$$
$$\text{(D2)}\ ((GH)\,(GK)) \vdash G\,((H)\,(K))$$
$$\text{(D3)}\ ((G)\,(H))\ ((G)\,(K)) \vdash ((G)\,(HK))$$
$$\text{(D4)}\ ((G)\,(HK)) \vdash ((G)\,(H))\ ((G)\,(K))$$

*Proof.* The derivations in $\mathfrak{S}_{qBa}$ are displayed as follows:

$$\begin{array}{cl} G\,((H)\,(K)) & \\ \hline G\,((GH)\,(K)) & (\text{Pit}) \\ \hline G\,((GH)\,(GK)) & (\text{Pit}) \\ \hline ((GH)\,(GK)) & (\text{Del}) \end{array}
\qquad
\begin{array}{cl} ((GH)\,(GK)) & \\ \hline ((GH)\,(GK))\ ((GH)\,(GK)) & (\text{Pit}) \\ \hline ((G)\,(G))\ ((H)\,(K)) & (\text{4 times Del}) \\ \hline ((G))\ ((H)\,(K)) & (\text{Pdeit}) \\ \hline G\,((H)\,(K)) & (\text{Inv2}) \end{array}$$





In the derivation of (D2), (Pdeit) is used as follows: the (Pdeit) is applicable to the graph $((G)^+\{(G)\})\;((H)\,(K))$ since the second occurrence of the partial graph $(G)$ can be deiterated from $(G)$ in the same area. Note that $(G)^+\{(G)\}$ is the partial graph of $((G)\,(G))$. Similarly, (Pdeit) is applicable in the derivation of (D3). This completes the proof. □

**Proposition 2.** *The following absorption and idempotency laws are derivable in* $\mathfrak{S}_{qBa}$*:*

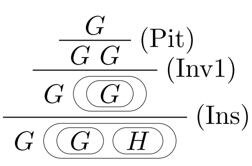
(Ab1) $G\,((G)\,(H)) \vdash G$ (Ab2) $G \vdash G\,((G)\,(H))$

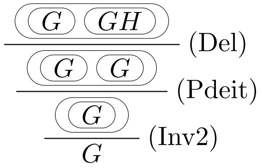
(Ab3) $((G)\,(GH)) \vdash G$ (Ab4) $G \vdash ((G)\,(GH))$

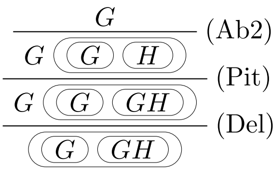
(Ide1) $GG \vdash G$ (Ide2) $G \vdash GG$

(Ide3) $((G)\,(G)) \vdash G$ (Ide4) $G \vdash ((G)\,(G))$

*Proof.* Clearly (Ab1) is obtained by (Del). The derivations of (Ab2), (Ab3) and (Ab4) are given as below:







Obviously (Ide1) and (Ide2) are obtained by (Del) and (Pit) respectively. (Ide3) is obtained by (Pdeit) and (Inv2). (Ide4) is obtained by (Inv1) and (Ins). □

**Proposition 3.** *The following distributive laws are derivable in* $\mathfrak{S}_{qBa}$*:*

(DM1) $(GH) \vdash (((G))\,((H)))$

(DM2) $(((G))\,((H))) \vdash (GH)$

(DM3) $(((G)\,(H))) \vdash (G)\,(H)$

(DM4) $(G)\,(H) \vdash (((G)\,(H)))$

*Proof.* (DM1) is obtain by twice (Inv1). (DM2) is obtained by twice (Inv2). (DM3) and (DM4) are obtained by (Inv2) and (Inv1) respectively. □

On Peirce's derivation of distributive laws in the Boolean algebra as well as in his classical Alpha graphs, see [10].

## 4 Soundness and Completeness

In this section, we prove the soundness and completeness of $\mathfrak{S}_{qBa}$ with respect to the variety $\mathsf{qBa}$.

**Definition 7.** *The translation $\tau\colon \mathcal{T} \to \mathfrak{G}_\alpha$ from the set of all terms to the set of all Alpha graphs is defined inductively as below:*

$$\tau(p) = p;\ \tau(\top) = \mathsf{SA};\ \tau(\bot) = \boxed{\phantom{x}}\ ;\ \tau({\sim}\phi) = \boxed{\tau(\phi)};\ \tau(\phi \wedge \psi) = \tau(\phi)\ \tau(\psi).$$

*For any non-empty finite multi-set $\Gamma = \{\phi_1, \dots, \phi_n\}$, let $\tau(\Gamma) = \tau(\phi_1)\dots\tau(\phi_n)$.*

*The translation $\rho\colon \mathfrak{G}_\alpha \to \mathcal{T}$ from the set of all Alpha graphs to the set of all terms is defined inductively as below:*

$$\rho(p) = p;\ \rho(\mathsf{SA}) = \top;\ \rho(\boxed{G}) = {\sim}\rho(G);\ \rho(G_1G_2) = \rho(G_1) \wedge \rho(G_2).$$

*For any graph $G$, let $\tau \circ \rho(G) = \tau(\rho(G))$.*

Let $\mathbb{A} = (\mathfrak{A}, \wedge, \vee, {\sim}, 0, 1)$ be a $\mathsf{qBa}$. For any assignment $\theta$ in $\mathbb{A}$ and an Alpha graph $G$, define $\theta(G) = \theta(\rho(G))$. We say that a graph $G$ is a *logical consequence* of $H$ with respect to $\mathsf{qBa}$, notation $G \models_{\mathsf{qBa}} H$, if $\theta(G) \leq \theta(H)$ for any assignment $\theta$ in any $\mathsf{qBa}$ $\mathbb{A}$. A graph $G$ is *equivalent* to $H$ with respect to $\mathsf{qBa}$, notation $G \equiv_{\mathsf{qBa}} H$, if $G \models_{\mathsf{qBa}} H$ and $H \models_{\mathsf{qBa}} G$.

**Lemma 1.** *For any graphs $G, H$ and $K$, the following hold in $\mathfrak{S}_{qBa}$:*

(1) *If $G \models_{\mathsf{qBa}} H$ and $H \models_{\mathsf{qBa}} K$, then $G \models_{\mathsf{qBa}} K$.*
(2) *If $G \models_{\mathsf{qBa}} H$, then $GK \models_{\mathsf{qBa}} HK$.*
(3) *If $G \models_{\mathsf{qBa}} H$, then $\boxed{H} \models_{\mathsf{qBa}} \boxed{G}$.*
(4) *If $G \models_{\mathsf{qBa}} H$, then $K^+\{G\} \models_{\mathsf{qBa}} K^+\{H\}$.*
(5) *If $G \models_{\mathsf{qBa}} H$, then $K^-\{H\} \models_{\mathsf{qBa}} K^-\{G\}$.*
(6) *If $G \equiv_{\mathsf{qBa}} H$, then $K\{G\} \equiv_{\mathsf{qBa}} K\{H\}$.*
(7) *$GH^+\{J\} \equiv_{\mathsf{qBa}} GH^+\{GJ\}$.*
(8) *$\boxed{\boxed{G}} \equiv_{\mathsf{qBa}} G$.*

*Proof.* (1)–(3) are obvious by definition. (4) and (5) are shown easily by simultaneous induction on the construction of $K$. (6) is easily shown by induction on the construction of $K$. Details are omitted here. For (7), we have $GJ \models_{\mathsf{qBa}} J$. By (4), $H^+\{GJ\} \models_{\mathsf{qBa}} H^+\{J\}$. Hence $GH^+\{GJ\} \models_{\mathsf{qBa}} GH^+\{J\}$. Conversely, it is easy to show that $GH^+\{J\} \models_{\mathsf{qBa}} GH^+\{GJ\}$ by induction on the construction of $H^+\{\ \}$. (8) holds because of the involution of $\sim$ in $\mathsf{qBa}$s. □

**Theorem 3 (Soundness).** *If $G \vdash_{\mathfrak{S}_{qBa}} H$, then $G \models_{\mathsf{qBa}} H$.*

*Proof.* Assume that $G \vdash_{\mathfrak{S}_{qBa}} H$. Let $H_0, \ldots, H_n$ be a derivation of $H$ from $G$ in $\mathfrak{S}_{qBa}$. We show that $G \models_{\mathsf{qBa}} H_i$ by induction on $i \leq n$. The case that $i = 0$ is obvious. Let $i > 0$. we have the following cases:

(1) $H_i$ is obtained from $H_{i-1}$ by (Del). Let $H_{i-1} = J^+\{K\}$ and $H_i = J^+\{\mathsf{SA}\}$. By induction hypothesis, $G \models_{\mathsf{qBa}} J^+\{K\}$. Clearly $K \models_{\mathsf{qBa}} \mathsf{SA}$. By Lemma 1 (1), $J^+\{K\} \models_{\mathsf{qBa}} J^+\{\mathsf{SA}\}$. Therefore $G \models_{\mathsf{qBa}} J^+\{\mathsf{SA}\}$.
(2) $H_i$ is obtained from $H_{i-1}$ by (Ins). Let $H_{i-1} = J^-\{K\}$ and $H_i = J^-\{KL\}$. By induction hypothesis, $G \models_{\mathsf{qBa}} J^-\{K\}$. Clearly $KL \vdash_{\mathfrak{S}_{qBa}} K$. By Lemma 1 (2), $J^-\{K\} \models_{\mathsf{qBa}} J^-\{KL\}$. Therefore $G \models_{\mathsf{qBa}} J^-\{KL\}$.
(3) $H_i$ is obtained from $H_{i-1}$ by (Pit), (Pdeit), (Inv1) or (Inv2). We obtain that $G \models_{\mathsf{qBa}} H_i$ by induction hypothesis and Lemma 1 (6)–(8).

**Lemma 2.** *For any graphs $G, H$ and $K$, the following hold in $\mathfrak{S}_{qBa}$:*

(1) *If $G \vdash_{\mathfrak{S}_{qBa}} H$ and $H \vdash_{\mathfrak{S}_{qBa}} K$, then $G \vdash_{\mathfrak{S}_{qBa}} K$.*
(2) *If $G \vdash_{\mathfrak{S}_{qBa}} H$, then $GK \vdash_{\mathfrak{S}_{qBa}} HK$.*
(3) *if $G \vdash_{\mathfrak{S}_{qBa}} H$, then $\boxed{H} \vdash_{\mathfrak{S}_{qBa}} \boxed{G}$ .*
(4) *If $G \vdash_{\mathfrak{S}_{qBa}} H$, then $K^+\{G\} \vdash_{\mathfrak{S}_{qBa}} K^+\{H\}$.*
(5) *If $G \vdash_{\mathfrak{S}_{qBa}} H$, then $K^-\{H\} \vdash_{\mathfrak{S}_{qBa}} K^-\{G\}$.*
(6) *If $G \equiv_{\mathfrak{S}_{qBa}} H$, then $K\{G\} \equiv_{\mathfrak{S}_{qBa}} K\{H\}$.*

*Proof.* (1) and (2) are obvious by the definition. For (3), assume that $G \vdash_{\mathfrak{S}_{qBa}} H$. Then there is a derivation $H_0, \ldots, H_n = H$. We show that $\boxed{H_i} \vdash_{\mathfrak{S}_{qBa}} \boxed{G}$ by induction on $i \leq n$. The case that $i = 0$ is obvious. Let $i > 0$. If $H_i$ is obtained from $H_{i-1}$ by (Pit), (Pdeit), (Inv1) or (Inv2), it is easy to obtain that $\boxed{H_i} \vdash_{\mathfrak{S}_{qBa}} \boxed{G}$ by induction hypothesis and the rule. Suppose that $H_i$ is obtained from $H_{i-1}$ by (Del). Let $H_i = J^+\{\mathsf{SA}\}$ and $H_{i-1} = J^+\{K\}$. By induction hypothesis, $\boxed{J^+\{K\}} \vdash_{\mathfrak{S}_{qBa}} \boxed{G}$. Clearly $\boxed{J^+\{K\}} \vdash_{\mathfrak{S}_{qBa}} \boxed{J^+\{\mathsf{SA}\}}$. Hence $\boxed{J^+\{\mathsf{SA}\}} \vdash_{\mathfrak{S}_{qBa}} \boxed{G}$. The case that $H_i$ is obtained from $H_{i-1}$ by (Ins) is shown similarly.

For (4) and (5), assume that $G \vdash_{\mathfrak{S}_{qBa}} H$. We prove that $K^+\{G\} \vdash_{\mathfrak{S}_{qBa}} K^+\{H\}$ and $K^-\{H\} \vdash_{\mathfrak{S}_{qBa}} K^-\{G\}$ by simultaneous induction on the construction of $K$. The case that $K$ is atomic is obvious. Assume that $K = K_1K_2$. Let $K^+\{\ \} = K_1K_2^+\{\ \}$. By induction hypothesis, $K_2^+\{G\} \vdash_{\mathfrak{S}_{qBa}} K_2^+\{H\}$. By (2), we obtain that $K_1K_2^+\{G\} \vdash_{\mathfrak{S}_{qBa}} K_1K_2^+\{H\}$. Let $K^-\{\ \} = K_1K_2^-\{\ \}$. By induction hypothesis, $K_2^-\{H\} \vdash_{\mathfrak{S}_{qBa}} K_2^-\{G\}$. By (2), we obtain that $K_1K_2^-\{H\} \vdash_{\mathfrak{S}_{qBa}} K_1K_2^-\{G\}$. Assume that $K\{\ \} = \boxed{J\{\ \}}$ . Then $K^+\{\ \} = \boxed{J^-\{\ \}}$ and $K^-\{\ \} = \boxed{J^+\{\ \}}$ . By induction hypothesis, $J^+\{G\} \vdash_{\mathfrak{S}_{qBa}} J^+\{H\}$ and $J^-\{H\} \vdash_{\mathfrak{S}_{qBa}} J^-\{G\}$. By (3), $\boxed{J^+\{H\}} \vdash_{\mathfrak{S}_{qBa}} \boxed{J^+\{G\}}$ and $\boxed{J^-\{G\}} \vdash_{\mathfrak{S}_{qBa}} \boxed{J^-\{H\}}$.

For (6), the proof proceeds by induction on the construction of $K$. The case that $K$ is atomic is obvious. Suppose that $K = K_1K_2\{\ \}$. By induction hypothesis, $K_2\{G\} \equiv_{\mathfrak{S}_{qBa}} K_2\{H\}$. By (2), $K_1K_2\{G\} \equiv_{\mathfrak{S}_{qBa}} K_1K_2\{H\}$. Suppose that $K = \boxed{J\{\ \}}$. By induction hypothesis, $J\{G\} \equiv_{\mathfrak{S}_{qBa}} J\{H\}$. By (3), $\boxed{J\{G\}} \equiv_{\mathfrak{S}_{qBa}} \boxed{J\{H\}}$ . □

**Lemma 3.** *If* $\mathsf{GqBa} \vdash \Gamma \Rightarrow \psi$*, then* $\tau(\Gamma) \vdash_{\mathfrak{S}_{qBa}} \tau(\psi)$*.*

*Proof.* The proof proceeds by induction on the height of a derivation of $\Gamma \Rightarrow \psi$ in $\mathsf{GqBa}$. Details are omitted here. Note that Propositions 1, 2, 3 and Lemma 2 are used.

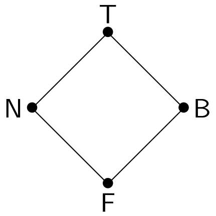
**Lemma 4.** *For any graph* $G$*,* $\tau \circ \rho(G) \equiv_{\mathfrak{S}_{qBa}} G$*.*

*Proof.* By induction on the construction of $G$. The case that $G$ is atomic is obvious. Let $G = G_1G_2$. Then $\tau \circ \rho(G) = G_1G_2$. Then $\tau \circ \rho(G) \equiv_{\mathfrak{S}_{qBa}} G$. Assume that $G = \boxed{H}$. Then $\tau \circ \rho(G) = \boxed{\tau \circ \rho(H)}$. By induction hypothesis, $\tau \circ \rho(H) \equiv_{\mathfrak{S}_{qBa}} H$. By Lemma 2 (3), $\boxed{\tau \circ \rho(H)} \equiv_{\mathfrak{S}_{qBa}} \boxed{H}$.

**Theorem 4 (Completeness).** *If* $G \models_{\mathsf{qBa}} H$*, then* $G \vdash_{\mathfrak{S}_{qBa}} H$*.*

*Proof.* Assume that $G \models_{\mathsf{qBa}} H$. Then $\rho(G) \models_{\mathsf{qBa}} \rho(H)$. By the completeness of $\mathsf{GqBa}$ with respect to $\mathsf{qBa}$, $\mathsf{GqBa} \vdash \rho(G) \Rightarrow \rho(H)$. By Lemma 3, $\tau \circ \rho(G) \vdash_{\mathfrak{S}_{qBa}} \tau \circ \rho(H)$. By Lemma 4, $G \vdash_{\mathfrak{S}_{qBa}} H$. $\square$

By the completeness theorem, one can show that $\mathsf{SA} \nvdash_{\mathfrak{S}_{qBa}} \boxed{\boxed{p}\,\boxed{\boxed{p}}}$ and $\mathsf{SA} \nvdash_{\mathfrak{S}_{qBa}} \boxed{p\,\boxed{p}}$. Consider the following four-valued distributive lattice with the unary operation $\sim$ [5]:



| $x$ | $\sim x$ |
|---|---|
| T | F |
| F | T |
| N | N |
| B | B |

Let $\theta(p) = \mathsf{N}$. Then $\theta(\boxed{p}) = \mathsf{N}$. Hence $\theta(\boxed{\boxed{p}\,\boxed{\boxed{p}}}) = \mathsf{N}$ and $\theta(\boxed{p\,\boxed{p}}) = \mathsf{N}$. But $\theta(\mathsf{SA}) = \mathsf{T}$. Hence $\mathsf{SA} \not\models_{\mathsf{qBa}} \boxed{\boxed{p}\,\boxed{\boxed{p}}}$ and $\mathsf{SA} \not\models_{\mathsf{qBa}} \boxed{p\,\boxed{p}}$. By the completeness theorem, $\mathsf{SA} \nvdash_{\mathfrak{S}_{qBa}} \boxed{\boxed{p}\,\boxed{\boxed{p}}}$ and $\mathsf{SA} \nvdash_{\mathfrak{S}_{qBa}} \boxed{p\,\boxed{p}}$.

## 5 Application and Discussion

Quasi-Boolean algebras are generalisations of Boolean algebra. A common structure for many-valued logics, including fuzzy logics and the Dunn–Belnap four-valued logic, quasi-Boolean algebras have found application in approximative reasoning and reasoning with inconsistent information.

We mention one area of interest for our proposed weak Alpha graphs. Since $\boxed{\phantom{p}}$ does not represent Boolean complementation, graphical languages that have been divested of that meaning tolerate inconsistencies and may be better suited for representing conflicting and vague information. We see the role of the $\mathsf{SA}$ indeed to be different in $\mathfrak{S}_{qBa}$ precisely in that in this logic we no longer assume the sheet to be the Sheet of *Assertions*.[5] That is, those who scribe and interpret graphs on the sheet need not bear responsibility for the truth of propositions. This in fact comes close to the interpretation Dunn and Belnap proposed for four-valued logic and the logic of first-degree entailments: such logics represent normative reasoning of how computers *should think* or *should behave* [5,7]. Machines, unlike minds, work with "told values" and not with truth-values *per se*. Slaves to syntactic manipulation, they do not question the propositions put to them, even if the information presented were inconsistent.

Accordingly, we could call the sheet in $\mathfrak{S}_{qBa}$, not the Sheet of Assertion but the Sheet of *Instruction*. Scribing the graphs on the sheet expresses not an act of assertion but a delivery of instruction. We can nevertheless go further and propose that the question of how computers 'should think' does not stop at the level of 'being instructed about' or 'being told' the propositions. What the graphical languages in general, and the expressions of $\mathfrak{S}_{qBa}$ in particular bring to the picture is precisely that their values are not 'told' but *shown*. Scribing them on the sheet is to show something rather than saying. When a computer is shown a graph $G$ of $\mathfrak{S}_{qBa}$, we show a computer the graph $G$ while we do not show $\textcircled{G}$, and when we show the graph $\textcircled{G}$ we do not show $G$. Moreover, since the Law of Non-Contradiction cannot be derived in this logic, we can show $G$ and $\textcircled{G}$. That is, we can bring both instances of such graphs to the "field of distinct vision"—a phrase used by Peirce to mean the range of the interpreter's attention (R 280, 1906)—of a sensory machine. Fourth, since the Law of Excluded Middle is not valid either, showing a graph $G$, including that of showing the blank SA, is not an expectation to commit computer to a truth-value regarding that graph. Showing a graph does not imply showing the *value* of that graph, and a machine could as well behave quite innocently at the presence of what might look like a contradiction in the classical language of EGs. As to the fourth case, recall also that the only axiom in classical EGs, namely that the blank sheet is a tautology which can always be added anywhere on the sheet juxtaposed with other graphs, is not present in $\mathfrak{S}_{qBa}$. Thus the common convention we find in Peirce's own theory, namely that "We always have a logical right to a blank sheet" (R 497, 1897), is not universal.

The present modification is only one suggestion among many others concerning the ways in which one could modify the significations of the basic geometrical and topological building blocks of these graphical languages to get new logics. For example, other generalisations of Boolean algebra are very well possible. We mention as examples the following three:

1. *Distributive lattices*, in which the SA does not represent anything at all and there are no ovals ◯. The only primitive operations are disjunction (represented as $n$-scrolls) and juxtaposition. Transformation rules lack insertion and iteration/deiteration and all areas are positive.
2. *Bounded distributive lattices*, which add the blank SA as $\top$ and ◯ as $\perp$.

---

[5] To the opposite direction, namely making it explicit that the sheet is precisely that of the assertions, see [3,4].

3. *Ockham algebras*, which take the ovals to be real *cuts*, that is, they are complementations such that an empty cut on the SA is a complement of that sheet, and a cut around the complement of the sheet is the sheet itself. Insertions on negative areas are not permitted, however.

Each of such generalisations can then become the desired structure for a graphical language in the large family of these supra-classical EGs.[6]

## 6 Concluding Remarks

Peirce introduced the Alpha part of the logic of Existential Graphs (EG) as a diagrammatic syntax and a graphical system corresponding to classical propositional logic. The logic of quasi-Boolean algebras is a logic below classical Alpha which we have used in the present paper as the algebraic structure for a new graphical system of weak Alpha graphs. It was shown to be sound and complete with respect to the class of all quasi-Boolean algebras. We also studied some implications of these modifications regarding the meaning of the basic components of EG, including the meaning of the SA, and proposed that these modifications can find applications in machine-related reasoning that use sensors to grasp information-based logical content. Thus various non-classical graphical logics are not only possible but important and come about with only minor modifications to the transformation rules of Peirce's original theory of EGs.

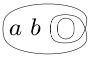
This overall idea is reflected in our proposed modification and its algebraic underpinnings. The modification is to have a *graphical language* that (i) can deal with such additional values detached from what the sheet naturally represents, (ii) is a system of deep inferences increasingly in demand for non-classical logics that represent and reason about inconsistent, finite and bounded computational processes, and (iii) preserves a conceptually and historically motivated approach that started off from Peirce's original insights.

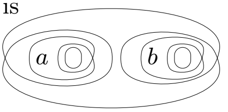
## References

1. Bellucci, F., Pietarinen, A.-V.: Existential graphs as an instrument of logical analysis: part I. Alpha. Rev. Symb. Logic **9**, 209–237 (2016)
2. Bellucci, F., Pietarinen, A.-V.: From Mitchell to Carus: 14 years of logical graphs in the making. Trans. Charles S. Peirce Soc. **52**(4), 539–575 (2017a)
3. Bellucci, F., Pietarinen, A.-V.: Assertion and denial: a contribution from logical notations. J. Appl. Logics **25**, 1–22 (2017b)

---

[6] Instead of a weakening of the classical Alpha to get the proposed quasi-Boolean graphs, we could strengthen the implication-free fragment of the intuitionistic system of Alpha [13] for the same effect. There the scroll  would only represent involution and there is no implication. A new rule is

From 

to 

.

4. Bellucci, F., Chiffi, D., Pietarinen, A.-V.: Assertive graphs. J. Appl. Non-Class. Logics (2018). https://doi.org/10.1080/11663081.2017.1418101
5. Belnap Jr., N.D.: A useful four-valued logic. In: Dunn, J.M., Epstein, G. (eds.) Modern Uses of Multiple-Valued Logic, pp. 5–37. Springer, Dordrecht (1977). https://doi.org/10.1007/978-94-010-1161-7_2
6. Brünnler, K.: Deep Inference and Symmetry in Classical Proofs. Logos, Berlin (2003)
7. Dunn, J.M.: A relational representation of quasi-Boolean algebras. Notre Dame J. Formal Logic **23**(4), 353–357 (1982)
8. Guglielmi, A.: Deep Inference. http://alessio.guglielmi.name/res/cos/
9. Ma, M., Fei, L.: Sequent calculi for semi-De Morgan and De Morgan algebras. Stud. Logica (2017). https://doi.org/10.1007/s11225-017-9755-7
10. Ma, M., Pietarinen, A.-V.: Peirce's sequent proofs of distributivity. In: Ghosh, S., Prasad, S. (eds.) ICLA 2017. LNCS, vol. 10119, pp. 168–182. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-662-54069-5_13
11. Ma, M., Pietarinen, A.-V.: Proof analysis of Peirce's alpha system of graphs. Stud. Logica **105**(3), 625–647 (2017)
12. Ma, M., Pietarinen, A.-V.: Gamma graph calculi for modal logics. Synthese (2017). https://doi.org/10.1007/s11229-017-1390-3
13. Ma, M., Pietarinen, A.-V.: A graphical deep inference system for intuitionistic logic. Log. Anal. (in press)
14. Oostra, A.: Los gráficos Alfa de Peirce aplicados a la lógica intuicionista. Cuad. Sist. Peirceana **2**, 25–60 (2010)
15. Peirce, C.S.: On an improvement in Boole's calculus of logic. Proc. Am. Acad. Arts Sci. **7**, 250–261 (1867)
16. Peirce, C.S.: On the algebra of logic. Am. J. Math. **3**(1), 15–57 (1880)
17. Peirce, C.S.: On the algebra of logic: a contribution to the philosophy of notation. Am. J. Math. **7**(2), 180–196 (1885)
18. Peirce, C.S.: Letter to Christine Ladd-Franklin, 9 November 1900. (R L 237)
19. Peirce, C.S.: Manuscripts in the Houghton Library of Harvard University, as identified by Richard Robin (1967). Cited as R followed by manuscript number
20. Pietarinen, A.-V.: Signs of Logic: Peircean Themes on the Philosophy of Language, Games, and Communication. Springer, Dordrecht (2006). https://doi.org/10.1007/1-4020-3729-5
21. Pietarinen, A.-V.: Is non-visual diagrammatic logic possible? In: Gerner, A. (ed.) Diagrammatology and Diagram Praxis. College Publications, London (2010)
22. Pietarinen, A.-V.: Exploring the Beta quadrant. Synthese **192**, 941–970 (2015)
23. Roberts, D.D.: The existential graphs and natural deduction. In: Studies in the Philosophy of C. S. Peirce, pp. 109–121. University of Massachusetts Press (1964)
24. Roberts, D.D.: The Existential Graphs of Charles S. Peirce. Mouton, The Hague (1973)
25. Shin, S.-J.: The Iconic Logic of Peirce's Graphs. MIT Press, Cambridge (2002)
26. Sowa, J.: Knowledge Representation. Brooks, Pacific Grove (2000)
27. Zeman, J.: The Graphical Logic of Charles S. Peirce. Ph.D. dissertation, University of Chicago (1964)

# Assertive and Existential Graphs: A Comparison

Ahti-Veikko Pietarinen[1,2](✉) and Daniele Chiffi[2]

[1] Nazarbayev University, Astana, Kazakhstan
ahti-veikko.pietarinen@ttu.ee
[2] Tallinn University of Technology, Tallinn, Estonia
chiffidaniele@gmail.com

**Abstract.** Peirce's graphical logic of Existential Graphs (EGs) has no specific sign for assertion, although the notion is used virtually everywhere in Peirce's logical theories. We outline the new system of Assertive Graphs (AGs) that makes the embedded notion of assertions in EGs explicit, and show how to inferentially transform AGs to a classical graphical logic CLAG, without having to introduce polarities explicitly. We compare the philosophy of notation of AGs to EGs, where the latter has polarities both in its intuitionistic and classical cases. Our comparison is framed with respect to three different representations of implication, namely as cuts, boxes and scrolls. We also identify three fundamental differences in the meaning of the Sheet of Assertion and compare those with Peirce's own proposed interpretation.

**Keywords:** Assertive Graphs · Existential Graphs
Intuitionistic graphs · Assertion · Polarity · Implication
Sheet of Assertion

## 1 Introduction

Various graphical languages to capture logic based on Peirce's method of Existential Graphs (EGs) have burgeoned in recent years. This paper provides a synopsis and a philosophical and notational comparison of three recent proposals: Assertive Graphs (AGs, [4]), intuitionistic EGs (the system GrIn of [11], cf. [12]) and classical Assertive Graphs (CLAG, [6]). We compare three major representational aspects of these languages: namely the notation that they employ to represent logical implication (the sign of conditional), the phenomenon of the polarity of areas that these graphs may or may not have, and the differences in the meaning of the Sheet of Assertion (SA) assumed in these graphical logics.

---

A.-V. Pietarinen—The work supported by the Estonian Research Council Personal Research Grant PUT 1305 (*Abduction in the Age of Fundamental Uncertainty*, PI Pietarinen), and Nazarbayev University Social Policy Grant 2018–2019.

D. Chiffi—The work supported by the Estonian Research Council Personal Research Grant PUT 1305 (*Abduction in the Age of Fundamental Uncertainty*, PI Pietarinen).

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 565–581, 2018.
https://doi.org/10.1007/978-3-319-91376-6_51

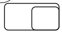
We focus only on propositional cases, which in Peirce's terms means the Alpha part of the theory of EGs. We assume familiarity with the basic features of his Alpha system.

Peirce's original proposal was classical in the sense that the logic of Alpha graphs satisfies bivalence and that all classically valid tautologies come out as true in it. Yet he laid considerable emphasis on the notion of *assertion* involved in the graphical method of logic. For example, in R 492 (1903) he writes that the SA is such that the universe that it represents is *definite* (i.e., the Principle of Non-Contradiction comes out as a tautology), *individual* (i.e., the Law of Excluded Middle comes out as a tautology), and *real.*[1] He explained the last point as "so that what is true and what false of it is independent of any judgment of man or men, unless it be that of the creator of the universe; in case this is fictive" (R 482). Any graph scribed on the sheet is in Peirce's view being asserted of some universe satisfying these three requirements. SA is a representation of that universe or universes, and an existential graph is a representation of a fact that exists in those universe(s).

Indeed assertion has enjoyed a wide and important role in a number of occasions in the development of modern logic. In addition to being an essential component of EGs, assertion has a distinguished inferential role in Frege's logic of judgments [7], and it has been used by Heyting in the explication of intuitionistic constants of logic [9]. Assertion has likewise been a major speech-act class analysed in illocutionary and pragmatic logics [5,17].

Yet the logic of assertion itself has not been a subject of investigation in the context of logical graphs and diagrams. A new diagrammatic system, called Assertive Graphs (AGs), was developed in [4] in order to provide a formal and philosophical account of the logic of assertions in the graphical context. AGs build upon many of the elements of EGs but also differ from them in three major respects: First, AGs take the assertion itself as the central constant of that logic. Second, the theory relies on intuitionistic rather than classical interpretation of logical constants. Third, the graphs of AGs need not distinguish between positive and negative areas of the graphs and hence the language does represent polarities.

It was shown in [6] that with an addition of one inferential rule, called Elimination of Coinciding Corners, intuitionistic logic of AGs can be converted into a classical graphical logic of assertions. The output is the system called CLAG, from classical assertive graphs. Significantly, the conversion is done without having to introduce polarities explicitly into the system. Since also intuitionistic variants of EGs have been proposed [11,12] which do rely on the representation of polarities, it now becomes topical to compare some of the key notational and philosophical aspects of these classical and intuitionistic versions of both EGs and AGs.

The present paper summarizes the basics of these three graphical logics and presents a notational comparison between them. The notion of assertion is congenial in all of them. Specifically, we will focus on three different graphical notations for implication that these logics make use of, namely by *cuts*  as in classical EGs, by *scrolls*  as in intuitionistic EGs (GrIn), and by *boxes*  as in both

---

[1] The reference R is to Peirce's manuscripts by the Robin number [15].

AGs and CLAG. We then discuss the *pros* and *cons* of dispensing with polarities in the graphical languages of AGs and CLAG. Third, we recount the role of the SA, as is remains as a central notion in both AG and EGs: The SA represents both the universe of discourse as well as provides the platform on which graphs are scribed as juxtaposed with each other and continuously connected by the blank of the sheet. The blank of the SA, we notice, is not empty, but it has different meanings in these three different notational systems: in EGs it means *all truths*, in AGs *all justified assertions*, and in intuitionistic EGs, the blank means *all transformations*.

## 2 Assertions in Graphs

We begin with a brief exposition of the theory of assertive graphs (AGs) [4,6].

### 2.1 Intuitionistic Assertive Graphs (AGs)

**Constructions.** The logic of AGs justifies intuitionistically valid principles. Its logical constants have a meaning that agrees with Heyting's explication of intuitionistic constants [9]. In short, an assertion is justified when it becomes possible to provide a *construction* (or a method of verification, demonstration, transformation etc.) that yields a proof of a proposition. Briefly, taking $P$ and $Q$ as propositions, (i) $P$ and $Q$ can be asserted iff both $P$ and $Q$ can be asserted, (ii) $P$ or $Q$ can be asserted iff at least one of the propositions $P, Q$ can be asserted, (iii) $\neg P$ can be asserted iff one possesses a construction which, from the supposition that a construction for $P$ were carried out, leads to a contradiction, (iv) the implication $P \rightarrow Q$ can be asserted iff one possesses a construction $R$ which, joined to any construction proving $P$ would effect a construction proving $Q$—that is, a proof of $P$, together with $R$, would form a proof of $Q$.

The notion of proof is an informal one. It may be, for instance, a valid and rigorous mathematical argument, or its skeletal idea. An assertion is justified when it becomes possible to possess a construction that leads to the proof of a proposition. This assertion-based explication of logical constants, together with the graphical and diagrammatic elements of EGs spreading the syntax into two-dimensions, provides an inspiration for the construction of AGs. The meaning of the SA might then concern *all justified assertions*: the justification condition for the act of assertion is conveyed by a construction that yields the proof for a propositional content of the graph instance scribed on the SA.

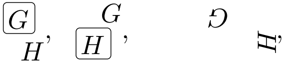
**Conventions and Rules of AGs.** We briefly describe the system of AGs. The expressions of the language of that system are graph-instances standing for assertions and their relations. We use $G, H, J$ etc. as names of propositions to stand for graph-instances. All graphs that are scribed on the SA are graph-instances. The fundamental conventions are listed first.

**Convention 1:** We always have a right to a blank SA.

**Convention 2:** We denote the assertion of a graph by scribing $\alpha$ on the SA enclosed within a *box*. Whenever $\alpha$ is the proposition $G$, then we can also scribe $\boxed{G}$, and *vice versa*. That is, rectangular boxes can be written around graphs and removed around them at will.

The box merely adds a *stress* and as such pertains to the paralinguistic or prosodic component of a language: one way of reading it would be to take it to draw our attention to its contents $G$, such as stating "Look, here's a $G$".

**Convention 3:** A juxtaposition is an assertion of graphs on the SA at two different positions of the sheet, such as $\boxed{G}\ \boxed{H}$. Graphs asserted at different positions of the SA are *independently* asserted. Since the conjunction of an assertion is equivalent to the assertion of its conjuncts, this is equivalent to $\boxed{G\ H}$. By virtue of these conventions, this is moreover equivalent also to any of $\boxed{G}\quad H$, $G\quad \boxed{H}$, $G\qquad\quad H$, $H\,G$, etc. Since SA is unordered in all directions, the following graphs are likewise equivalent to the above:



, *etc.*

**Convention 4:** Two juxtaposed graphs conceived not independently but *alternatively* asserted are connected by a *thin line* with a bar crossing it, such as $\boxed{G}\!+\!\boxed{H}$.

**Convention 5:** The nesting of boxes with the inner and outer boxes connected by sharing two adjacent sides represents an *implication* between two assertions, notated by $\boxed{G\,\boxed{H}}$.

This is *not* equivalent to $\boxed{G\,\boxed{H}}$, which by Convention 2 expresses a conjunction of two assertions.

**Convention 6:** An absurdity is represented by *the blot*: $\bullet$.[2]

**Convention 7:** Nothing else is a convention.

The well-formed graphs of the language of AGs are defined inductively along the lines of these conventions. We skip the details and refer to [4,6]. The logic is then defined by the set of its graphical axioms and the rules of transformation on the graphs of the language of AGs.

**Axioms of AGs.** There are three axioms.

**Axiom I.** The blank SA: $\qquad,\ \boxed{\phantom{HH}}$

**Axiom II.** (Any graph implies a blank): $\boxed{H\,\boxed{\phantom{H}}}$

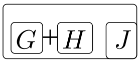
**Axiom III.** (*Ex falso*): $\boxed{\bullet\,\boxed{H}}$.

**Rules of Transformation.** There are altogether nine transformation rules in the logic of AGs.

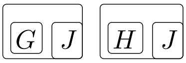
---

[2] The blot is assumed to blacken the entire area within which it occurs. Since it is impractical to show this by actually blackening large blank areas, a heavy bullet is used instead.

**(As/Am)** The Rule of Antecedent Separation/Antecedent Merging:



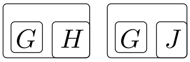
if and only if



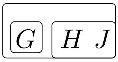
.

That is, the disjunction of the antecedents of a conditional can be split into juxtaposition. Conversely, any two conditionals with the same consequent can be merged into a conditional with the disjunction of the antecedents.

**(Cm/Cs)** Consequent Merging/Consequent Separation:



if and only if

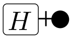


.

That is, the consequents of two conditionals with identical antecedents can be merged into the consequent of a conditional. Conversely, juxtaposed consequents in a conditional can be split into the juxtaposition of two conditionals.

**(DC)** Rules of Disjunct Contradiction:

if and only if

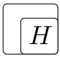


.

That is, any graph is equivalent to its disjunction with the blot.

**(CR)** The Cornering Rule:

$H$ if and only if



.

That is, any graph is equivalent to the conditional having the same graph as a consequent with a blank antecedent.

**(It/Deit)** Iteration/Deiteration Rule: For this rule, we introduce the notion of the context of graphs. A context is of the form $K\{\}$, where a single slot $\{\}$ is the empty context. Let $K\{H\}$ be the graph obtained from $K\{\}$ by substituting $H$ for the slot. The two rules are:

**(It)** If a graph $G$ occurs on the SA or anywhere in the nest of graphs, it may be scribed on any area (not part of $G$) which is contained by $\{G\}$:

If $K\{GH\{J\}\}$ then $K\{GH\{GJ\}\}$.

The converse of (It) is deiteration (Deit):

**(Deit)** If $K\{GH\{GJ\}\}$ then $K\{GH\{J\}\}$.

**(CE)** Conjunction Elimination:

If

then

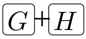
.

That is, from the assertion of $G$ and $H$ it is possible to derive the assertion of one of them.

**(DI)** Disjunction Introduction:

If

then



.

That is, from the assertion of $G$ it is possible to derive the assertion of $G$ or the assertion of $H$.

**(IA)** Insertion in the Antecedent:

$$\text{If}\quad K\{\boxed{G\boxed{H}}\}\quad\text{then}\quad K\{\boxed{G\ J\boxed{H}}\}.$$

The rule (IA) is applied to the contexts $K\{\}$ of graphs. The restriction of (IA) is to conditionals whose immediate context in $K$ (that is, the area on which the conditional is placed) is not that of the antecedent of another conditional. That is, in any unoccupied position in the area of the antecedent of the conditional which itself does not reside in an antecedent of a conditional of its immediate context, it is possible to insert any graph.

**(DC)** Deletion from the Consequent:

$$\text{If}\quad \boxed{G\boxed{H\ J}}\quad\text{then}\quad \boxed{G\boxed{H}}.$$

That is, one can delete any graph from a consequent of a conditional.

With these rules, one can derive all intuitionistic validities, and one cannot derive the Law of Excluded Middle (LEM), for example [4,6].

### 2.2 Classical Assertive Graphs (clag)

A classical variant of ags, called clag, is obtained by introducing the following rule of Eliminating Coinciding Corners **(CC)** for *absurdum*:

**(CC)** Eliminating Coinciding Corners:

$$\text{If}\quad \boxed{\boxed{G\boxed{\bullet}}\boxed{\bullet}}\quad\text{then}\quad G.$$

That is, if there are no intermediate graphs (other than the blank) between two corners each occupied by nothing but a blot, then these two corners cancel each other out.

What remains after the application of this rule is $\boxed{\boxed{G}}$, which is the same as the graph $G$. The rationale for the rule **(CC)** is a natural graphical and spatial property: in the absence of any intermediate graphs occupying the intermediate space between two corners occupied by nothing but the blots, then these two corners would annihilate each other out (imagine the amorphous blots welding into each other). In clag, we can now derive Peirce's Law, which is equivalent to the LEM and to the elimination of a double negation $\neg\neg G \rightarrow G$, as well as all the other classically valid principles.

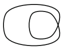
## 3 Intuitionistic Existential Graphs (GrIn)

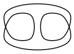
Our third example of a graphical system that introduces certain new notational innovations is an intuitionistic modification of Peirce's original egs, called GrIn [11]. After a brief introduction to GrIn, we compared its notation with the key elements of ags and egs, namely the representation of implication, polarity, and the meaning of the SA.

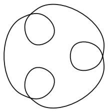
### 3.1 The Language of Intuitionistic Graphs

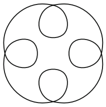
Intuitionistic EGs are constructed on the SA by their diagrammatic syntax which is inductively defined from atoms by the operation of *juxtaposition* and by the following $n$-ary scrolls for finite $n > 0$:

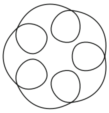

Nullary scroll

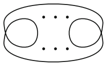

Unary scroll


Binary scroll

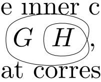
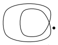

Ternary scroll

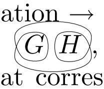
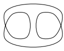

Quadruple scroll


Quintuple scroll


$n$-tuple scroll, $n > 1$

There are three types of atomic graphs: (1) a denumerable set of *propositional letters* Prop, (2) the Sheet of Assertion (SA) and (3) the *nullary scroll* , which has nothing but a blank in its interior. The unary scroll has one *outer close* and one *inner close*. For $n \geq 2$, the $n$-ary scroll has one outer close and $n$ inner closes.

The graph obtained by scribing a graph $G$ at the outer close and a graph $H$ at the inner close of the unary scroll is called the *implicational graph*. It is notated as , and it has an antecedent $G$ and a consequent $H$. The graphical sign that corresponds to the intuitionistic implication $\rightarrow$ is thus .

The *binary scroll* of $G$ and $H$, notation , is interpreted as the operation of disjunction. The graphical sign that corresponds to the intuitionistic disjunction commonly symbolized as $\ovee$ is thus .

The graph $G\,H$ is called the *juxtaposition* of $G$ and $H$ on the SA, and is interpreted as the operation of conjunction.

### 3.2 Rules

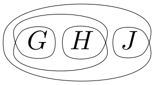
Given this language as informally described above, the graphical system GrIn consists of four axioms and six rules of transformation. The RULES OF DELETION, INSERTION and ITERATION/DEITERATION are as in classical EG. The first two are irreversible and they appeal to the polarity of the area being positive and negative, respectively. Rules of iteration and deiteration are reversible rules. Obviously there is no rule of double cuts. Instead, there are the following additional rules:

**(BA1/BA2)** THE RULES OF THE BLANK ANTECEDENT:

$G\{H\}$ if and only if $G\{$  $\}$.

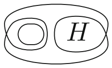
That is, any partial graph (that is, a sub-graph) of a graph can be replaced by  and *vice versa*.

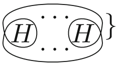
**(AmR/AmS)** THE ANTECEDENT MERGING/SPLITTING RULES:

$K\{$   $\}$ if and only if $K\{$  $\}$.

That is, the antecedents of two scrolls with the same consequent can be merged into a double-scroll. Conversely, two graphs of a double-scroll in the antecedent of a scroll can split into the juxtaposition of two scrolls.

**(CD1–4)** THE RULES OF DISJUNCTION:

$G\{H\}$ if and only if $G\{$  $\}$, and

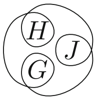
$G\{H\}$ if and only if $G\{$  $\}$.

The first two of these rules (**(CD1)** and **(CD2)**) mean that any partial graph $H$ of a graph $G$ can be replaced by disjuncting a pseudo-graph  to it and *vice versa*. The second pair (**(CD3)** and **(CD4)**) mean that any partial graph $H$ of a graph $G$ can be replaced by any number of equivalent inloops and *vice versa*.

The above systems have the obvious properties of *colligation*, *commutativity* and *associativity*. They follow from the properties of the space of the SA, which is an open-compact, isotropic manifold. Given the non-linear syntax of these graphs, such properties need no separate statement at the level of rules. In fact we would not be able to state such rules in the "diagrammatic syntax" (R 485, 1897-8; 515, L 376, 1911) of these graphical logics. Rather, we *show* them. Since the SA is continuous, and since by Convention 3 a co-occurrence of two propositions is an assertion of them both, the colligation merely means that two independently asserted graphs on the same area are juxtaposed, which is a conjunction. No separate rule of conjunction expresses colligation, which just means that partial graphs are co-located in the same area of the sheet.

Commutativity is shown in GrIn by the equivalence of graphs such as , , . Also  and  are exactly the same graph, for example.

Likewise, in a 3-scroll



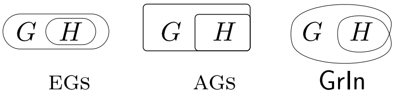
the inloops can be at any position in the area of the outloop, that is, they can traverse past each other without overlapping. Commutativity of intuitionistic disjunction thus means that the order and the position on the outloop in which the inner closes of an $n$-ary scroll occur is immaterial. Inloops have no designated position for their intersection points on the outloops. Associativity follows from commutativity of the inloops of the double scroll for disjunction, which is a primitive sign. The resulting system GrIn is sound and complete with respect to the class of all Heyting algebras [11].

## 4 A Notational Comparison of Conditionals, Polarities and the Sheets of Assertion

In this section, we compare these three graphical logics, EGs, AGs (and CLAGs) and GrIn, with respect to three characteristic features: the representation of *conditional assertions* in terms of the notations for the sign of implication, the role of *polarities* in the representation of these graphical languages, and the signification of the sheet of assertion SA in them. We also contrast these notions with Peirce's original insights, as much of those derive from his writings that have so far remained unpublished.

### 4.1 Comparing Three Graphical Representations of Conditionals

The signs of implication ($G \to H$) are shown in these graphical logics in three different ways. The first, notation with nested cuts, is the implication in classical EGs, the second, the box-notation, is the implication in AGs, and the third, the scroll, is the notions for implication in intuitionistic EGs:



Notational differences between these three representations of implication are important. Classical implication with nested cuts has zero intersection points between the cuts. Intuitionistic implication in GrIn has precisely one intersection point; it is where the inloop and the outloop connect. The sign of implication in the box-notation for assertive graphs has two adjacent sides between the two nested rectangulars that coincide, that is, the intersection consists of the interval of the two adjacent sides of the inner box which is strictly smaller than the outer box.

These differences are all material and distinctive characteristics of the respective representations. The intuitionistic meaning of the implication is as an operation that transforms the construction for its antecedent to that of the construction for its consequent. Unlike what is the case in Peirce's material conditional *de inesse* in his EGs, intuitionistic scroll shows the connection between the antecedent and consequent in terms of the non-separation between the two constructions: at least one intersection point is shared between the boundaries that define the areas of the antecedent (the area of the outloop) and the areas of the consequent (the area of the inloop).

Peirce's view on conditionals *de inesse* was that in graphs of the theory of EGs such as ,

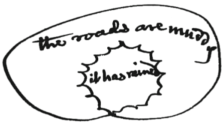
> [n]either A nor C is positively asserted; and therefore neither can be written on the sheet of assertion. They must be cut off from that sheet. Moreover, they must be separated from one another, and in such a manner that they shall not appear as reciprocally related in any one way. They must therefore be in separate enclosures of a different character in one from the other, but connected. (R 482, alternative draft)

Peirce observes that in material conditionals, the notation may not exhibit any connection between the antecedent and the consequent, as the conditional could just as well be taken to be a negation of two conjunctive assertions, and conjunctions are juxtapositions *independently asserted* on the SA.

Typically Peirce would use both the nested-cut and the scroll notation interchangeably, assuming that the inloop of the scroll could be detached from the outloop at will. "The node", he claims in R 482, "or point of intersection, has no particular significance". He took both the nested-cut and the scroll notations as insignificant typological variations that can be interpreted in exactly the same way. He assumes conditionals to be *de inesse*, that bivalence holds, and that the LEM as well as other rules of EGs that yield classical logic are valid, and thus that the logic that one gets would correspond to the Boolean algebra [10].

Yet this assimilation of the two notations of nested cuts and the scrolls, and thus the detachment requirement for conditional signs, was not without exceptions in Peirce's writings. His earliest manuscripts on EG, written in late 1896 (R 481, 482, 488), all begin the exposition of the graphical language with the conditional drawn as a scroll, never as two nested and detached ovals. There are also some further, collateral reasons that we can learn from his philosophy of logic that indicate that some of Peirce's ideas were not far removed from an intuitionistic interpretation of logical constants. For example, he graphically derived the sign of negation from the sign of implication [1]. Nearly all of his main presentations concerning EGs interpret negation as validating the double-cut rule, however, so that both introduction and elimination of double negations come out as valid.

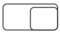
This said, the scroll  as an intuitionistic implication has an interesting interpretation originating from Peirce's remarks concerning the method of how these graphical forms are to be conceived and how they are to be transformed to some other graphical forms. Both in his algebraic and graphical logics, the conditional is the primary logical constant because it is the only primitive relation that mirrors illation. Negation is defined by the conditional, not *vice versa* (R 481). For this idea to work, the inner scroll which contains a pseudo-graph in its enclosure would need to disappear—not the two boundaries of the scroll, the outloop and the inloop. This those boundaries must not be exactly of the same quality and form. In R 481 Peirce indeed draws the inloop of the scroll with a qualitatively different, saw-edged shape, thus distinguishing it from the outloop which he retains as a simple closed curve. Those signs of implication that we find in R 481 are given in the graph below, in Peirce's hand for "If the roads are muddy, then it has rained":



Later in 1906 (R 292, 295, S-30) he draws this qualitative difference between the representations of the inloop and the outloop in words and calls the outloop "the Wall" and the inloop "the Fence". We could take this as a sign of an anticipation of the intuitionistic idea that the interpretation of the conditional should concern assertions of the *method of construing a method*:[3] there has to be a method (or a proof, transformation, verification etc.) that applies, as soon as the method is applied to the antecedent of the conditional, also to the consequent of that conditional. The metaphor of the outer close being bound by "the Wall" from the outside captures the idea that the method according to which a graph becomes entitled to be asserted on the outer close will at once also guarantee the method of asserting a graph on the inner close. (Take the latter being surrounded by a construction weaker than walls.) Once you have penetrated to the area of the antecedent, you will be able to penetrate to the area of the consequent as well. Movement is from the graphs permitted to be asserted on the antecedent area, to the graphs permitted to be asserted on the consequent area.

Interpretations of graphical logical constants by a close notational examination are not mere figures of speech or grounded on typographical minutiae. In them, we hear Peirce's pre-intuitionistic conception of mathematics, namely that mathematics is the exact study of *mental creations*.[4] Peirce took mental and imaginative creations and constructions literally as objects that are to be best analyzed in their diagrammatic outfit. They are not abstract objects. As soon as diagrams can be given precise logical interpretations and the associated calculi of rules of transformations (namely having logical graphs scribed upon the "Phemic sheet" which is endowed with the property of them being at once asserted; see R 855, 669, 292, S-30), one is prepared to suggest that these systems fulfill the purpose of analysing the nature of mathematical constructions, and ultimately the nature of mathematical reasoning.

## 4.2 Comparing Polarity

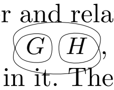
The assertive and intuitionistic characters of conditionals are clearly two different things. They can be distinguished by digging deeper into the nature of the two graphical representations. There is more than one singular intersection point, namely the interval intersected in . The remaining boundary of the inner box separates the areas of the antecedent and that of the consequent. This means that information from the area outside of the conditional into the antecedent area of that conditional (say by the application of the rule of iteration) can traverse either across the overlapping or the non-overlapping boundaries, that is, either passing or not passing through the area of the antecedent.

---

[3] This reflects Peirce's own turn of phrase from 1885: "But I cannot doubt that others, if they will take up the subject, will succeed in giving the notation a form in which it will be highly useful in mathematical work. I even hope that what I have done may prove a first step toward the resolution of one of the main problems of logic, that of producing *a method for the discovery of methods in mathematics*" (added emphasis).

[4] A historical tidbit is that the first to notice that Peirce's conception of negation in logic might mean that the LEM would not hold was Gerrit Mannoury, Brouwer's supervisor. Peirce was surely keen to limit the applicability of the LEM to propositions that are *determinate*.

In classical EGs, there would always be a polarity switch when traversing across the boundary of the cut, from the negative to the positive area or *vice versa*. But in the box notation of AGs, the rectangular shapes do not give rise to polarities. They are not cuts. Everything in assertive graphs rests on the sheet of assertion. Boxes do not cut their interiors, nor are they punctuation marks like the cuts in EGs. Only implicitly do the cornered boxed introduce the distinction between what lies on the antecedent and what on the consequent areas (cf. also the rule (IA)). The absence of cuts means that nothing is removed from the sheet and nothing rests anywhere else than on the sheet. Thus the precise location by which the information traverses in these graphs *endoporeutically*[5] is immaterial. The absence of polarities distinguishes AGs also from intuitionistic EGs, as in the latter the scrolls do demarcate two different qualities of areas, and crossing the boundary means switching the polarity.

Another and related notational difference of note is that in $n$-scrolls ($n \geq 2$), such as in , the area between the inloops is inactive and nothing may be scribed in it. The language of AGs make this manifest by taking disjunctions as primitives in which disjuncts are graphs which are not *independently* but *alternatively* asserted. This is notated by having them connected by lines, so that no spatial area between is present at all, such as in $\boxed{G}{+}\boxed{H}$ or $G{+}\boxed{H}{+}J$ etc.

The polarities can as well be interpreted, as Peirce at least since 1906 often did, in terms of the difference between the two *sides* of the SA, namely in terms of the *verso* and the *recto* leafs of the sheet. In GrIn, this means that the area of the outloop of the scroll is an area that resides on the reverse, *recto* side of the leaf of the sheet, while the area of its inloop marks an area on the leaf that obtains on the *verso* side.

Here no two graphs of which one is scribed on the antecedent and the other on the consequent area are independently asserted, since the justification of an intuitionistic implication is that it is *conditional upon* the method of transformation which is applied to its antecedent area (such as insertions), such that the method in question would also suffice to verify the justification of the assertion of the graph in the area of the antecedent. That the former is cut off by "the Wall" as its boundary means that there is a reversal of the side of the leaf upon which the graph is scribed, while the latter being cut off by "the Fence" as its boundary means that the leaf is returned from its reversed position to match the quality of the original sheet (as erasures are permitted on both sides of the leaf). Recall that these boundaries do not mean negations: their meaning concerns the intuitionistic interpretation of implicational signs.

In terms of its inferential system, representation of logic that retains polarities as EGs also retains the irreversible rules of insertion and erasure. These are the only rules that depend on the presence of boundaries (such as walls and fences) that also are markers to switch polarities. This is what contradictory negations do in EGs; in a weaker sense they may be other operations such as

---

[5] That is, from the outside-in direction, from the context of the conditional inside conditional forms, see R 292b, 293, 300, 515, 650, 669.

involutions in so far as they preserve the duality. When as in AGs and CLAGs the boundaries do not mark polarity, insertion and erasure rules may be supplanted by (restricted) insertions to the antecedents of a conditional and by erasures from the consequents of those conditionals.

None of the boundaries in these notations directly characterize negative operations. Negation is defined in terms of an implication of the pseudo-graph (the empty cut or the blot). That the boundary is not the cut is the common feature that distinguishes these graphical logics from Peirce's original method of EGs. However, this does not mean that Peirce would have missed to consider the alternative: since his 1880 paper on the algebra of logic, he had represented negation as an implication of the falsity (such as $A \rightarrow 0$, where 0 means $\perp$). By omitting just one of the axioms from his 1885 algebra, namely the fifth icon (Peirce's Law), one would get an intuitionistic kernel of that logic.

## 4.3 Comparing the Meaning of the Sheets

What is the meaning of the SA in these different cases? Peirce's original idea of the sheet is that nothing can be scribed on the sheet which is not at once also asserted. Scribing expresses an *act* of asserting, and the sheet is endowed of the property that whatever comes to transpire upon it as the result of the scribing is also at once asserted. The pseudo-graph, which signals absurdity, is not an assertion as a boundary is not something that can be scribed.

We made this idea explicit in the logic of AGs. The act of assertion is a logical constant of that system. One of the virtues of this approach is that the sign of assertion is an *embedded* sign in the notion of the sheet. We can thus avoid typical problems that proposals that have explicit, or *ad hoc*, signs of assertion would accrue [2]. There are no polarities and no incisions by negations, and hence all areas are areas on the SA. Consequently, the implication is strictly speaking not a hypothetical conditional, because also the area of the antecedent rests on the SA. This is so also in the classical rendering of CLAG.

In GrIn, in contrast, the conditional, as represented by the scroll  is hypothetical. Its antecedent does not rest on the SA, nor does the area of its inloop. But those areas are not cut off from the sheet as there are no cuts. We take the scroll to have a natural intuitionistic interpretation in which constructions associated with conditionals are graph transformations.

Spatial transformations are indeed central in graphical logics. What the diagrammatic syntax of these logics boils down to is that areas of graphs are constituted by a *continuum* of positions. Rules of transformation can be applied at arbitrary positions. This has the effect that the inferential component of these logics is that of a *deep inference*. (See [8] for a review of the current literature.) GrIn is a fundamental logic of illative transformations. There are, for example, many more permissible transformations in GrIn than in EGs.

The rule of blank antecedents (BA1/BA2), for example, has an important intuitionistic interpretation. As asserting a graph $G$ on the SA may be thought of as an equivalent to "there is a transformation of $G$", then in intuitionistic graphs, unlike in standard Alpha graphs, asserting a graph by scribing it on the SA is not an existence claim concerning that proposition. Scribing a graph means *committal to the producibility of a method for its transformation.* Hence the scroll around $G$, , means that the space of transformations that occupy the antecedent area, as indicated by the blank, implies the graph $G$.

Therefore intuitionistic graphs are strictly speaking not species of *existential* graphs. Assertions that are made by scribing graphs on intuitionistically conceived sheets of assertion are assertions neither of a property nor an identity of anything that exists in the universe of discourse represented by the sheet. They are affirmations that *one can find* that what is asserted by some suitable method, such as there being a series of transformations that would bring about what the assertion expresses. The presence of transformations affirms the presence of a demonstration or verification of those objects of the graphs that assertion-makers scribe on the SA.[6]

Peirce explained "existential" in EGs to refer to facts or characters that exist in the mutually recognized universe when we scribe graphs on the SA and thus assert them as signifying those facts and characters:

> [I]ts fundamental symbol expresses the relation of existence. I speak of existence as a relation, because it consists in the occurrence of a nature among a collection of individual objects of experience,—not necessarily all actually experienced, but all destined to be experienced, could the experience be rounded out to completion. (R 485, 1897–8).

Intuitionistic graphs of the kind defined in GrIn express this positive relation as the graph's relation to the possibility of its illative transformations, its verifications or proofs. These graphs do not symbolize that relation as existence. The sheet upon which intuitionistic graphs are scribed is indexically connected to the universe of discourse which consists of those illative transformations as verificatory methods that pertain to the common body of experiences between those who undertake to discourse upon the graphs and to whom the meaning of the SA is assumed to be well understood, namely the utterers and the interpreters. Peirce often termed them the (imaginary) "graphist" and the "interpreter".[7]

---

[6] Similar remarks hold on other logical constants of GrIn besides the scroll. In intuitionistic Beta graphs the phenomenon that the blank of the continuous sheet is that of the space of all transformations becomes even more pronounced, since the line of identity could be interpreted as signalling an identity of proofs.

[7] "[I]f we take a piece of blank paper, and form the resolve to write upon it some part of what we think *about some real* or *imaginary condition of things*, then ... the whole sheet having been devoted to that purpose exclusively, by the common understanding called of the *graphist* (as the person who makes assertions by "scribing",—that is, by writing, drawing, or otherwise putting—on the sheet so devoted is to be called), and the *interpreter* (i.e. the person to whose understanding the graphist addresses the assertions that he scribes on the sheet), the graphist is at liberty to scribe any assertion on the sheet that he may be disposed to assert" (R 678).

Again, Peirce is looking for innovative meanings to the SA:

> [T]he sheet of assertion is a graph-instance, and its meaning is the meaning of all the standing permissions of the system, together with anything else that may be which is, in all cases well understood between the graphist and the interpreter. (R 280, 1904, alternative drafts)

The signification of the SA is the totality of what the standing permissions (the set of conventions plus the set of illative transformations) of the system allow it to stand for, including the mutually agreed interpretation of the graphical language. Those permissions include all graphs transformations scriptible upon the sheet. When the permissions change, the meaning of the SA changes accordingly. Likewise, by restricting the totality of the meaning of the standing permissions to the meaning of illative transformations which are topological constructions, one can endow the sheet with new significations such as intuitionistic ones.[8]

## 5 Conclusion

We have compared three approaches to graphical logics with respect to their representation and meaning of the sign of implication, polarity, and the sheet of assertion. We remarked on main differences these logics have with respect to Peirce's original method of EGs, and showed that the representation of polarity is not a necessary feature of neither intuitionistic nor classical graphical logic of assertions. This approach might pave the way towards graphical analyses of assertions where the choice of the method need not depends on differences of whether the underlying logic is intuitionistic or whether it subscribes to certain 'classical' rules such as Peirce's Law.

Our other finding was that the signification of the SA in intuitionistic logic is to represent all transformations. A closer look at Peirce's motivations in developing the graphical method of logic reveals that he chose the scroll, as the sign of implication, to be the fundamental logical constant because it mirrors the process of illation, and logical reasoning is about illative inferential relations [1]. In the intuitionistic case, this motivation comes out even stronger than in Peirce's own, 'classical' EGs, because the scroll embodies the idea later stated in Brouwer–Heyting–Kolmogorov interpretation as a construction that transforms the proofs of its antecedent into those of the consequent. The scroll itself is that construction and the illative rules involving those constructions codify the transformations. In intuitionistic graphs, SA is the representation of all such possible transformations. The blank sheet is not empty as it implicitly contains all the scrolls, just as the classical SA implicitly contains all the cuts that can be brought, as Peirce once put it, to the "field of distinct vision" or to the "mental

---

[8] There is a *continuum* of intermediate logics and as logical graphs they have not yet been studied. One would expect changes in the meaning of the SA to be an important indicator of logical differences between intermediate logics.

vision of [the] attention" (R 280, 1906) of the graphist and the interpreter who scribe and interpret these graphs.[9]

We noted in the introduction that this mutually recognized universe which the SA represents is in Peirce's view both *definite* (LNC: "so that no assertion can be both true and false of it"), *individual* (LEM: "so that any assertion is either true or false of it") and *real* ("so that what is true and what is false of it is independent of any judgment of man or men, unless it be that of the creator of the universe; in case this is fictive") (R 492, 1903). In intuitionistic graphs, we preserve the definiteness but are compelled to give up the property of individuality, though we do so without evoking any third truth value. If the universe is not definite, inconsistency-tolerant graphs may arise. Moreover, if the methods of transformation themselves are taken to govern judgments of those who undertake to discourse upon graphs, then we might also be compelled to give up the reality of the universe of discourse. Peirce's gadget was to take there to be also the "grapheus", who "creates the fictitious universe", "determines its characters" and "authorizes the assertions" of the graphist (R 492, R S-28). We leave any further examination of such wider philosophical repercussions of these non-classical graphical methods as the topic of another paper.

## References

1. Bellucci, F., Pietarinen, A.-V.: Existential graphs as an instrument of logical analysis: part I. Alpha. Rev. Symb. Logic **9**, 209–237 (2016)
2. Bellucci, F., Pietarinen, A.-V.: From Mitchell to Carus: 14 years of logical graphs in the making. Trans. Charles S. Peirce Soc. **52**(4), 539–575 (2017a)
3. Bellucci, F., Pietarinen, A.-V.: Assertion and denial: a contribution from logical notations. J. Appl. Logics **25**, 1–22 (2017b)
4. Bellucci, F., Chiffi, D., Pietarinen, A.-V.: Assertive graphs. J. Appl. Non-Class. Logics (2018). https://doi.org/10.1080/11663081.2017.1418101
5. Carrara, M., Chiffi, D., De Florio, C.: Assertions and hypotheses: a logical framework for their opposition relations. Logic J. IGPL **25**(2), 131–144 (2017)
6. Chiffi, D., Pietarinen, A-V.: On the Logical Philosophy of Assertive Graphs, preprint
7. Frege, G.: Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens. Louis Nebert, Halle a. S. (1879)
8. Guglielmi, A.: Deep Inference. http://alessio.guglielmi.name/res/cos/
9. Heyting, A.: Intuitionism: An Introduction. North Holland, Amsterdam (1956)
10. Ma, M., Pietarinen, A.-V.: Proof analysis of Peirce's alpha system of graphs. Stud. Logica. **105**(3), 625–647 (2017)
11. Ma, M., Pietarinen, A.-V.: A graphical deep inference system for intuitionistic logic. Log. Anal., in press

---

[9] "[S]ince a scroll both of whose closes are empty asserts nothing, it is to be imagined that there is an abundant store of empty scrolls on a part of the sheet that is out of sight, whence one of them can be brought into view whenever desired" (R 669, 1910). Here Peirce speaks of the scrolls as double cuts. This convention implies the double-cut rule. By taking the "abundant store of empty scrolls" to refer to intuitionistic domains we get a different semantics for implication.

12. Oostra, A.: Los gráficos Alfa de Peirce aplicados a la lógica intuicionista. Cuad. Sist. Peir. **2**, 25–60 (2010)
13. Peirce, C.S.: On the algebra of logic. Am. J. Math. **3**(1), 15–57 (1880)
14. Peirce, C.S.: On the algebra of logic: a contribution to the philosophy of notation. Am. J. Math. **7**(2), 180–196 (1885)
15. Peirce, C.S.: Manuscripts in the Houghton Library of Harvard University. Cited as R or L followed by manuscript or letter number (1967)
16. Roberts, D.D.: The Existential Graphs of C.S. Peirce. Mouton, The Hague (1973)
17. Searle, J.R., Vanderveken, D.: Foundations of Illocutionary Logic. Cambridge University Press, Cambridge (1985)

# On the Transformation Rules of Erasure and Insertion in the Beta Part of Peirce's Existential Graphs

Shigeyuki Atarashi(✉)

Department of Philosophy, Doshisha University,
601 Imadegwa-Karasuma, Kyoto 602-8580, Japan
`satarash@mail.doshisha.ac.jp`

**Abstract.** The aim of this paper is to show the symmetric organization of the transformation rules for erasing or inserting graphs in Existential Graphs, which Peirce constructed as the diagrammatical system of logic that uses graphical apparatus instead of formulae of standard formal logic. Many researchers have overlooked the symmetry of Peirce's initial layout of the transformation rules of erasure and insertion. In this essay, I will symmetrically rearrange the rules of erasure and insertion on the basis of the accurate understandings of Peirce's statements.

## 1 Diagrammatical Representations

Existential Graphs is divided into three parts; the Alpha part, the Beta part, and the Gamma part. Roughly speaking, the Alpha part is propositional logic, the Beta part corresponds to first-order predicate logic, and the Gamma part might be taken as modal logic. I will focus on the Beta part, because I think that we do not have a more precise grasp of Peirce's project in the part yet. In the standard notation of formal logic, "Something is an F and a G" is expressed in the following manner; $\exists x(Fx \wedge Gx)$. Existential Graphs represents it by connecting "F" and "G" with a heavy line, as the left-hand graph in Fig. 1 shows.

Existential Graphs calls these diagrammatical representations *graphs*. Peirce calls the heavy line that lies between "F" and "G" a *line of identity*, which asserts the identity of the individuals denoted by its two extremities [1, 4.443]. Peirce also uses a closed curve termed a *sep* or a *cut*, the effect of which is to deny the proposition represented by the entire graph within it [1, 4.402]. Peirce calls the area within a cut its *close* [1, 4.437]. By enclosing with a cut "G" of the left-hand graph in Fig. 1, we obtain the middle graph in Fig. 1 signifying the sentence, "Something is an F and not a G." If we enclose that entire graph with a cut, then we arrive at the graph meaning that there is nothing that is an F and not a G, as displayed by the right-hand graph in Fig. 1. The assertion of this graph conforms to the proposition, "Every F is a G," or, $\forall x(Fx \rightarrow Gx)$.

Peirce stipulates two basic rules for transforming graphs, one of which permits us to scribe a new graph in a designated close of a given graph, and is

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 582–589, 2018.
https://doi.org/10.1007/978-3-319-91376-6_52

**Fig. 1.** "Some" & "Every"

called *insertion* [1, 4.505]. For example, in the left-hand graph in Fig. 2, which is simplified here for brief explanation, insertion enables us to scribe "G" next to "F" in the close formed by the cut, so that we obtain the middle graph in Fig. 2. By insertion, we may also connect "F" to "G" with a heavy line, as the right-hand graph in Fig. 2 shows. This graph affirms that there is not anything that is at once an F and a G. (I will give their accurate expressions in Sect. 3.)

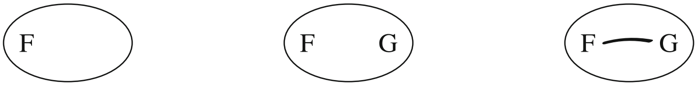


**Fig. 2.** Insertion

The other operation is *erasure* by which we may remove a graph from a designated close of a given graph. For instance, in the left-hand graph in Fig. 3, we are allowed to delete the heavy line lying between "F" and "G" by erasure. Thus, this operation leads us from the left-hand graph in Fig. 3 to the middle graph in Fig. 2. In the latter graph, by erasure, we may obliterate "G" from the juxtaposition of "F" and "G." Through these transformations, we conclude that something is an F (the right-hand graph in Fig. 3) from the premise that something is an F and a G (the left-hand graph in Fig. 3).

**Fig. 3.** Erasure

Many researchers have referred to Peirce's Existential Graphs from their own standpoints. In particular, Roberts made Peirce's complicated descriptions of the transformation rules much clearer [2]. However, Roberts' interpretations contain some problems, which have prevented us from observing a symmetry of Peirce's arrangement of the transformation rules of erasure and insertion in the Beta part. As Shin points out, a kind of symmetry is built in Peirce's systems of Existential Graphs [3, 81]. According to Shin, it interferes with a more efficacious development of Existential Graphs [3, 81]. However, I think that we can reconstruct a more systematic configuration of the rules on the basis of Peirce's original design of them. This is only feasible by an exact reading of the explanations that Peirce offers for the graphs and the rules in the Beta part.

## 2 Roberts' Reading

Roberts' reading of the transformation rules includes difficulties. The crucial problem is Roberts' notion of a line of identity. Roberts decides to adopt the conception that any line of identity can cross cuts [2, 50]. However, Peirce clearly asserts that "a line of identity is a partial graph; and as a graph it cannot cross a sep" [1, 4.499] Thus, in Fig. 4, Peirce considers the heavy line to be a series of two heavy lines abutting upon one another at the cut, and calls that series a *ligature* [1, 4.499]. In Peirce's terms, the line of identity in Fig. 2 is a ligature composed of two heavy lines; one of which is outside the cut and extended to the point of intersection on it; the other is inside it and joined to that point. Roberts' approach to Existential Graphs does not choose this notion of ligatures.

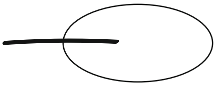


**Fig. 4.** Ligature

Roberts introduces the following rule: any evenly enclosed graph and any evenly enclosed portion of a ligature may be erased [2, 56]. If a graph is evenly enclosed, then it is enclosed by an even number of cuts. Roberts gives two examples to illustrate applications of this rule [2, 56]. In the first graph from the left in Fig. 2, by use of this rule, the portion outside the cut of the ligature may be erased, for it is enclosed with no cut, so that we obtain the second graph. The portion inside the inner cut of the ligature of the third graph is enclosed with two cuts, and we may transform it into the fourth graph by erasing that portion. Therefore, Roberts regards these applications as one and the same operation of erasure. (The graphs in Fig. 5 are simplified, and their accurate expressions will be given in Sect. 3.)

To the contrary, Peirce considers one of the two operations as a different transformation from the other. Peirce gives the following stipulation about one of them: "It permits any ligature, where evenly enclosed, to be severed from the inside of the sep immediately enclosing that evenly enclosed portion of it" [1, 4.505]. What type of operation is permitted by this rule? First, it refers to an evenly enclosed portion of a ligature. Second, a cut immediately encloses that portion of the ligature, which, therefore, lies inside the cut. In Fig. 5, the first and third graphs from the left have evenly enclosed portions of the ligatures. However, in the first graph, the evenly enclosed portion of the ligature lies outside the cut, while that of the third lies inside the cut. Hence, Peirce's above rule enables us to erase the evenly enclosed portion of the ligature of the third graph, so that we obtain the fourth graph. The rule cannot be applied to the evenly enclosed portion of the ligature of the first graph.

Peirce also introduces the following rule: "This rule permits any ligature, where evenly enclosed, to be severed" [1, 4.505]. It might seem to conform to the

**Fig. 5.** Examples of erasure

rule of erasure formulated by Roberts. We have observed that Peirce divides a ligature into a heavy line inside a cut and one outside it and that he gives the transformation rule for erasing the evenly enclosed portion inside the inner cut of the ligature of the third graph from the left in Fig. 5. Therefore, it should be assumed that this rule is formulated to erase the evenly enclosed portion outside a cut of a ligature. According to Peirce, points on a cut lie outside the close of the cut, and are treated as if they were away from it [1, 4.450]. In the first graph, the point of intersection on the cut at which two heavy lines outside and inside the cut are linked to each other is placed outside the close of the cut. Thus, the portion outside the cut of the ligature of the first graph from the left in Fig. 5 stands in the same situation as the ligature of the left-hand graph in Fig. 3. Both of them are deleted by the rule of erasure that Peirce prescribes with no reference to any cut.

## 3 Dots

This investigation reveals that the function of a heavy line consists in coupling one point with another. Peirce states that heavy lines have “the effect of joining dots on the sep to dots outside and inside of it” [1, 4.449]. Peirce uses the word “dot” instead of “point.” For example, in the third graph from the left in Fig. 5, the portion inside the outer cut and outside the inner cut of the ligature links the dot next to “F” with that of intersection on the inner cut, and the portion inside the inner cut links the dot next to “G” with it. The ligature does not directly connect “F” and “G” in Fig. 5. Perice calls “F” and “G” *spots*, and identifies them as those predicates with blanks, which we shape by eliminating their subjects from propositions of the form of subject-predicate [1, 4.441], for instance, “is a logician.” This blank being filled with a proper name, that predicate is turned into a complete proposition; “Peirce is a logician.” Peirce calls each blank of a spot a *hook*, and states: “A spot with a dot at each hook shall be a graph” [1, 4.441]. Hence, the diagrammatical representation for the sentence, ”Something is an F,” is the first graph from the left in Fig. 6, and the second graph stands for the sentence, “Something is an F and a G.”

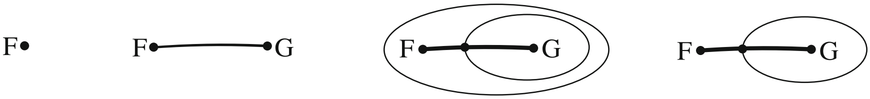


**Fig. 6.** Hook dots & cut dots

We may call a dot with which each hook of a spot is filled a *hook dot.* A dot on a cut may be called a *cut dot.* By using cut dots, we draw the third graph from the left in Fig. 6 as diagrammatically signifying the sentence, “Whatever is an F is a G,” and the fourth graph means the sentence, “Something is an F and not a G.” As we have seen, a line of identity or a heavy line does not cross any cut. In both of the third and the fourth graphs, each ligature is composed of the two heavy lines outside and inside the cut. In the third graph, the evenly enclosed portion inside the inner cut of the ligature may be erased inwards. In the fourth graph, since the non-enclosed portion of the ligature lies outside the cut, it may be erased outwards from the cut. We may term the former erasure *disjoin* and the latter *separation.* This classification divides the erasures in even enclosures into two kinds; *inward* and *outward.*

Basically, the rule of erasure is applied to the evenly enclosed portions of ligatures. In addition, Peirce puts forward the following rule: “It permits any ligature to be retracted from the outside of any sep in the same enclosure on which the ligature has an extremity” [1, 4.505]. According to this, we may erase that portion, either evenly or oddly enclosed, of a ligature that outwards extends to a dot on and stops at the cut which is placed in the same close as that portion. This erasure is about each ligature of the left-hand and the right-hand graphs in Fig. 7. We have already introduced the rule for erasing the ligature of the latter graph as separation, because any cut dot is coordinate with a dot outside the cut. Therefore, we should assume that the Peirce’s above statement sheds light, especially on the other kind of erasure that indicates the outward erasure in an odd enclosure. The rule allows us to erase the oddly enclosed ligature of the left-hand graph in Fig. 7 and transform it into the graph composed of a pair of cuts in whose inner close “G” with its hook dot lies and in whose annulus, in the area between the outer cut and inner cut, “F” with its hook dot lies without any heavy line, as the middle graph in Fig. 7 shows.

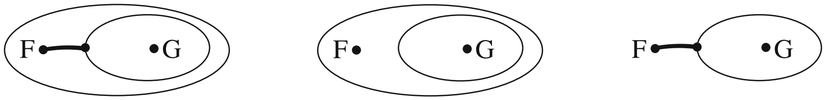


**Fig. 7.** Retraction

These considerations offer us a new conception of the way to organize the rules of erasure and insertion that Peirce originally gave. We can adjust the rules of erasure and insertion from two viewpoints; *outward* or *inward*; *even* or *odd* erasure, so that the eight rules can be symmetrically set out with their names:

1. The inward/outward erasure in an even enclosure: *disjoin/separation.*
2. The inward/outward insertion in an odd enclosure: *join/connection.*
3. The inward/outward erasure in an odd enclosure: *disjoint/retraction.*
4. The inward/outward insertion in an even enclosure: *joint/extension.*

## 4 Symmetric Arrangement

Now I illustrate what types of transformation these rules permit, based upon Peirce's original descriptions. Peirce declares about the rules of erasure and insertion as follows:

> This rule permits any ligature, where evenly enclosed, to be severed, and any two ligatures, oddly enclosed in the same seps, to be joined. It permits a branch with a loose end to be added to or retracted from any line of identity. It permits any ligature, where evenly enclosed, to be severed from the inside of the sep immediately enclosing that evenly enclosed portion of it, and to be extended to a vacant point of any sep in the same enclosure. It permits any ligature to be joined to the inside of the sep immediately enclosing that oddly enclosed portion of it, and to be retracted from the outside of any sep in the same enclosure on which the ligature has an extremity. [1, 4.505]

This passage is to be partitioned into seven sentences. (1) "This rule permits any ligature, where evenly enclosed, to be severed[.]" As we have seen in Figs. 5 and 6, this is separation. (2) It permits "any two ligatures, oddly enclosed in the same cuts, to be joined." This means the transformation of the left-hand graph in Fig. 8 into the right-hand graph, in which transformation the hook dot of "F" and the cut dot on the inner cut are oddly enclosed with the outer cut and the ligature links them outwards from the inner cut. Therefore, according to the above list, this transformation can be identified as connection.

**Fig. 8.** Connection

(3) "It permits a branch with a loose end to be added to or retracted from any line of identity." I have to clarify what a branch and a loose end stand for to explain the permits of this rule. I will address this later. (4) "It permits any ligature, where evenly enclosed, to be severed from the inside of the sep immediately enclosing that evenly enclosed portion of it[.]" In Figs. 5 and 6, it has been already demonstrated that this is disjoin.

(5) It permits any ligature, where evenly enclosed, "to be extended to a vacant point of any sep in the same enclosure." This operation shapes the left-hand graph in Fig. 9 into the middle graph. In the former graph, the hook dot of "F" and any vacant dot on the cut are in the same close as that cut. The ligature is permitted to reach the vacant cut dot from the hook dot of "F" outside the cut. The list of the rules tell us that it is to be characterized by extension. (6) "It permits any ligature to be joined to the inside of the sep immediately enclosing

that oddly enclosed portion of it[.]" By the rule, the right-hand graph in Fig. 9 is fashioned from the left-hand graph. In the latter graph, the hook dot of "G" is oddly enclosed, and the ligature is extended from the hook dot of "G" to a dot on the cut inwards. This is join. (7) It permits any ligature "to be retracted from the outside of any sep in the same enclosure on which the ligature has an extremity." This transformation is confirmed as retraction in Fig. 7.

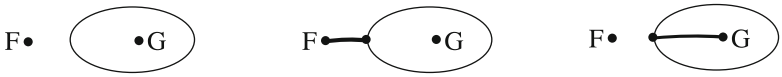


**Fig. 9.** Extension & join

Rule (3) remains to be examined: "It permits a branch with a loose end to be added to or retracted from any line of identity." According to Peirce, a branching ligature is a diagrammatical representation "signifying the identity of the three individuals" [1, 4.446]. A heavy line has its two extremities and designates the identity of the two individuals denoted by its two ends. Since a branching ligature denotes the identity of the three individuals, it has three extremities, as the first graph from the left in Fig. 10 shows.

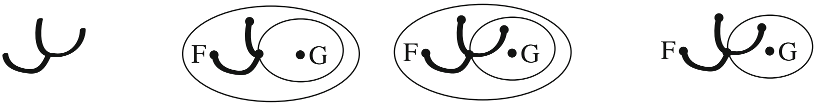


**Fig. 10.** Branches

A loose end is an extremity of a heavy line which links up with neither a hook nor a cut. A dot is put on a loose end, which dot may be called *a loose end dot*. As Peirce states, a heavy line with a loose end can be extended or retracted in one and the same close without being connected to any hook or cut [1, 4.502]. Hence, rule (3) authorizes us to eliminate from or add to a ligature a branch with a loose end dot in a different close, since rule (3) imposes no restriction on the erasure or the insertion of a branch with a loose end dot. The two rules remain to be explained; joint and disjoint. Joint is the inward addition to a ligature of a branch with a loose end dot within an evenly enclosed cut and permits us to transform the second graph from the left in Fig. 10 into the third graph. Disjoint is the inward retraction from a branching ligature of its branch with a loose end dot within an oddly enclosed cut. In the fourth graph from the left in Fig. 10, the heavy line branching inwards to the cut may be erased by this rule.

## 5 Conclusion

Peirce's way of formulating the rules of erasure and insertion in the Beta part of Existential Graphs covers all of the eight types into which I have classified them in the two pairs of terms; *outwards* and *inwards*, *evenly* and *oddly*. The notion that a line of identity crosses a cut precludes us from detecting the original symmetry which Peirce built in his prescriptions of them. To avoid this difficult, we need to pay much more attention to Peirce's basic conception that any line of identity cannot cross a cut, and to select the way of interpreting a line of identity as a ligature that links hook dots, sep dots, or loose end dots.

## References

1. Peirce, C.S.: Collected Papers of Charles Sanders Peirce. The Belknap Press of Harvard University Press, Cambridge (1974)
2. Roberts, D.D.: The Existential Graphs of Charles S. Peirce. Mouton, The Hague (1973)
3. Shin, S.-J.: The Iconic Logic of Peirce's Graphs. A Bradford Book, Cambridge (2003)

# Existential Graphs as a Basis for Structural Reasoning

Javier Legris(✉)

Instituto Interdisciplinario de Economía Política de Buenos Aires,
CONICET-Universidad de Buenos Aires,
Av. Córdoba 2122, C1120 AAQ Buenos Aires, Argentina
jlegris@retina.ar

**Abstract.** On the verge of the 20th century, Charles S. Peirce was convinced that his Existential Graphs were the *best* form of presenting every deductive argument. Between 1900 and 1909, Peirce chose the *scroll* as a basic sign in his Alpha system for Existential Graphs. According to a recent paper by Francesco Bellucci and Ahti-Veikko Pietarinen, the reason for this choice lies mainly in the *non-analyzable* nature of the *scroll*: *Only one sign* expresses the basic notion of *illation*. In this paper, some analogies between this early version of the Alpha system and *Structural Reasoning* (in the sense of Kosta Došen and Peter Schröder-Heister) are explored. From these analogies, it will be claimed that the system Alpha based on the *scroll* can be used as an accurate framework for (i) constructing basic structural deductions and (ii) accomplishing a diagrammatic interpretation of logical constants of First-Order Language. Moreover, EGs show cognitive advantages with respect to sequent systems. In this paper, the basic conception is outlined in an informal way, without making an exposition of the technical details.

**Keywords:** Diagrammatic reasoning · Existential graphs · Structural reasoning

## 1 Introduction: Peirce's Iconic Conception of Deduction

Charles Sanders Peirce (1839–1914) belongs to the "grounding fathers" of mathematical logic. He began his work in logic in the algebraic tradition of mathematical logic stemming from Boole. However, due to *philosophical* reasons, he became dissatisfied with the algebraic notation and he started developing at the end of the 19th century a *diagrammatical* system for logic: his *Existential Graphs* (EGs). As it is known, he regarded them as his masterpiece, his *chef d'oeuvre*, in logic, and in a letter to William James he characterizes them as the *logic of the future* (see the classic book [1], p. 11).

In fact, all mathematics was diagrammatic for Peirce, including mathematical proof. As he stressed in papers posthumously published in volume IV of *New Elements of Mathematics*, mathematical proof is a process of transformation of diagrams by showing their logical structure. Since diagrams are icons, a proof has an *iconic* function with respect to deduction. In Peirce's theory of signs, icons are characterized not only as being similar to their objects, but also as being signs that provide information through their observation and manipulation. This idea is already sketched in Peirce's seminal paper on algebra of logic from 1885 (see [2] 3.363 and 5.165).

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 590–597, 2018.
https://doi.org/10.1007/978-3-319-91376-6_53

In short, Peirce conceived his EGs as the *best* formulation of logical systems, according to both his *iconic* conception of logic and mathematics and his idea of an "exact logic" (*i.e*., mathematical logic). Peirce formulated different systems for the EGs that have been intensively studied and expanded in recent time on the basis of examining unpublished manuscripts. His *Alpha* system corresponds to propositional logic, the *Beta* system corresponds to Predicate Logic with identity and the *Gamma* system aims at modal and Higher-Order Logic.

The main underlying assumptions of this paper can be summarized as follows: (1) diagrams can be used as an intuitive way to express the nature of deduction (with interesting cognitive properties) and as a mathematical tool; (2) the theory for formulating logic systems on the basis of EGs evolves from a coherent philosophical perspective (i.e. Peirce's iconic conception of deduction), so that there is a clear correspondence between logical theory and philosophical theory; (3) the philosophical theory does not make deduction and logical concepts dependent on ordinary language, so that logical form is not a linguistic one. On this basis, this paper attempts to explore some fruitful analogies between Peirce's EGs and *Structural Reasoning* in the sense conceived by Došen and Schröder-Heister (see [3, 4]).

## 2 Peirce's Existential Graphs and the *Scroll*

In the first versions of the Alpha system for propositional logic, the basic sign was the *scroll*: a sole continuous line, giving rise to two attached or connected ellipses, one inside the other, both drawn in the *sheet of assertion* (see, *v.g*., Peirce MS 450, p. 14, quoted in [1], p. 34). The resort to the scroll relates this interpretation to Peirce's notion of *illation*, represented in the algebra of logic by the "craw foot" sign (–<) (see v.g. [2], 3.66, 3.139 and 3.175). Peirce also observed:

> "I thought I ought to take the general form of argument as the basal form of composition of signs in my diagrammatization; and this necessarily took the form of a 'scroll,' that is a curved line without contrary flexure and returning" ([2] 4.564).

The resulting diagram was the following:

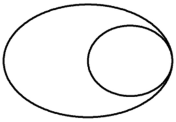


**Fig. 1.** Scroll. Example 1.

Of course, the same diagram is obtained if Fig. 1 is rotated like in Fig. 2.

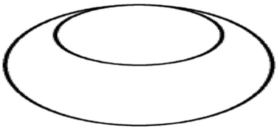


**Fig. 2.** Scroll. Example 2.

So, it should be clear that it is not the case of a one-dimensional notation in disguise.

The inner line of the scroll was designated as the *inner loop* of the scroll, the outer line being the *outer loop*. Any graph outside the inner loop and inside the scope of the outer loop was *hypothetical*, and any graph inside the inner loop *depended* on the hypothesis. Hence, a sentence of the form ‘Under the hypothesis A, B holds’ (or an explicit conditional sentence like ‘If A, then B’, where A and B are sentences) was represented in the Alpha System of EGs as (Fig. 3).

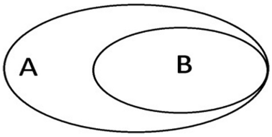


**Fig. 3.** The scroll as illative graph.

So, the scroll could be regarded as an *illative graph*.

It is easy to see that the procedure to “inscribe” the outer loop of the scroll can be iterated, so that a scroll can include several nested outer loops, as in the following Fig. 4, where the hypothetical role of A and B with relation to C is clear, C being implied by both hypothesis.

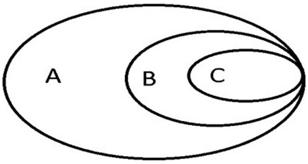


**Fig. 4.** Scroll with nested loops.

It must be noted that a *scroll* with more than one inner loops remains *a sole continuous line*.

However, this representation becomes troublesome in the formulation of rules of transformation for graphs in the Alpha system and in the proof of classical equivalences. The problems arise quite clearly in cases involving negation. For example, in classical logic, a sentence of the form ‘If A, then B’ is equivalent to ‘It is not the case that A and not B’. In later versions of EGs, Peirce represented negation with ‘cuts’ in the sheet of assertion, that is, with closed lines like the following (Fig. 5):

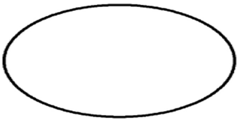


**Fig. 5.** The cut.

Now, the diagrammatic representation of ‘It is not the case that A and not B’ should be

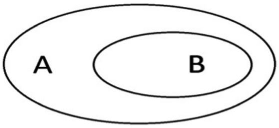


**Fig. 6.** The conditional as two cuts.

From a diagrammatic point of view, this representation is not the same as Fig. 2. The inner loop of Fig. 6 is not in contact with (or attached to) the outer loop. This fact had the consequence that Peirce, finally, represented the conditional using two nested cuts, as in Fig. 6, and not through a continuous line. On this basis, Fig. 6 represented both ‘If A, then B’ and ‘It is not the case that A and not B’ showing their logical equivalence. Both propositional forms are used in ordinary language to express the same logical structure. Therefore, there is in EGs a (implicit) *diagrammatic rule* concerning the *contact* of two lines. In fact, it comes to be irrelevant that the inner loop is in contact with the outer loop.

## 3 Structural Reasoning

The diagrams resulting from applying the *scroll* of EGs correspond to *structural deductions*. Kosta Došen pointed out that “the term ‘structural’ […] should be understood in the sense this word has in Gentzen’s sequent-systems” ([3], p. 364). From the point of view of structural reasoning, logic begins with structural deductions. According to Peter Schroeder-Heister,

> “By *structural reasoning* we mean reasoning in a sequent style system using structural rules only. Structural rules do not refer to the internal compositions of formulas by means of logical connectives or quantifiers but only affect the way formulas appear within sequents” ([4], p. 246).

This perspective stems from Gerhard Gentzen’s original ideas of in order to analyze deduction developed in his groundbreaking paper [5]. A sequent sign ‘⇒’ could be characterized as a *one-dimensional* or *linear* form of the scroll in Fig. 1, and the graph in Fig. 2 would correspond to the sequent

$$A \Rightarrow B,$$

Where A and B are, normally, sentences of First-Order Language (FOL). In this paper, only *singular* sequents of the general form

$$A_1, \ldots, A_n \Rightarrow C$$

are taken into account. Notwithstanding, it is possible to correlate the scroll with multiple sequents as they were originally used by Gentzen, that is, sequents of the form

$$A_1, \ldots, A_n \Rightarrow B_1, \ldots, B_m.$$

In fact, Peirce conceived scrolls with *multiple inner loops* as in Fig. 7 (see [2] 4.457).

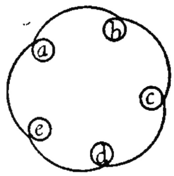


**Fig. 7.** Multiple inner loops.

(where the lower case letters stand for sentences). So, the scroll can be generalized in two senses: (a) scrolls containing embedded inner loops (see Fig. 4) or (b) scrolls containing many inner loops from an only one outer loop (see Fig. 7). In the inner loops, the sentences (represented by Peirce through lower case letters) should be understood disyunctively, like the formulas in the *succedens* of Gentzen's multiple sequents.

Gentzen's Structural rules, like thinning, contraction, interchange or cut (see [6], Sect. 3) have their counterparts in particular cases of the rules of *insertion*, *erasure* and *iteration* for the Alpha system (see [1] p. 40 *passim* and [2]. 4.425 ff.). Thinning in singular sequents corresponds to insertion in odd: Any graph may be scribed on any oddly enclosed area. Contraction is understood as a *special case* of deiteration: Any graph *in the outer loop* whose occurrence could be the result of iteration can be erased. Interchange corresponds to the fact that an order (absent in EGs) is set up in the system of sequents. The cut rule corresponds roughly to the transitivity of the scroll, a property that required a proof in the Alpha system.

An essential feature of the EG is their *analytic* nature. They are the *continuation* of Peirce's previous algebraic study of logic but carried out more accurately. Through his contributions both in algebra and in diagrammatic logic, Peirce tried "to dissect the operations of inference into as many steps as possible" (Peirce [2], 4.424). Hence, "exact" or mathematical logic had mainly an analytical role. In 1893 he wrote that the aim of logic was "to analyze reasoning and see what it consists in", and in relation to his exposition of Existential Graphs, he argued that it is "the business of logic to be analysis and theory of reasoning, but not the practice of it" ([2], 2.532 and 4.134).

Furthermore, this formulation of EGs provides the *right* analysis of logical concepts. The reason for that rests on Peirce's special idea of analysis as *uniqueness of decomposition*. This conclusion was recently reached by Francesco Bellucci and Ahti-Veikko Pietarinen on the basis of some passages from Peirce's unpublished manuscripts. According to Peirce, the system of Existential Graphs provides "the only method by which all connections of relatives can be expressed by a single sign", as "the System of Existential Graphs recognizes but one mode of combination of ideas" (MS 482, 1897 and MS 490, 1906, see [6], 211). The *scroll* is the *sole* sign that *cannot be decomposed*, and this is shown by the actions performed in constructing the sign, by inscribing "a curved line without contrary flexure and returning", in Peirce's own words. This is a pure "diagrammatological" fact, using the expression coined by Stjernfelt (see [7]). In the case of structural reasoning, a similar feature can be determined.

## 4 A Diagrammatic Interpretation of Logical Constants

Logical constants of FOL can be interpreted in the assertion sheet by means of existential graphs. For example, a way of characterizing negation as cut results from Peirce's scroll and his idea of continuity in diagrams. Under the condition that nothing is inscribed within the inner loop (it is a *pseudograph*, see [2], 4.454 f. and [1], p. 36), the scroll can be shrunk continuously, so that it is finally transformed into a cut, as illustrated in Fig. 8.

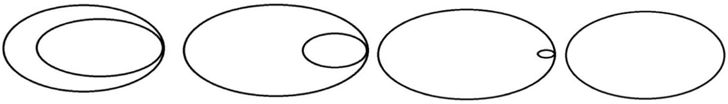
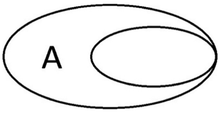
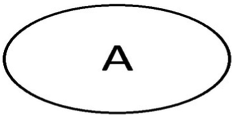


**Fig. 8.** The shrinking scroll and the pseudograph.

Hence, the graph



is ultimately transformed into the graph



(In fact, Peirce represented the pseudograph as a black spot entirely filling the loop that "may be drawn invisibly small" [2] 4.455).

This characterization can be compared with (and, furthermore, could also be seen as a diagrammatic proof of) the derivation of the sequent

$$\Rightarrow \neg \mathbf{A}$$

from the sequent

$$\mathbf{A} \Rightarrow$$

From this comparison, the sign ¬ of FOL in ¬**A** is interpreted by saying that **A** plays, in this sentence, the same role that it plays in the scroll when it is in the outer loop and the inner loop is a pseudograph. Hence, an *interpretation* in the EG of the logical constants of FOL could be devised. The procedure is, to some extent, analogous to the elucidation of logical constants by means of sequents.

A special case of this approach is the idea of analyzing logical constants in terms of structural deductions developed by Došen, where logical constants point out structural features of deductions (in the sense of features that are purely schematic, see [3], 365 f.). According to Došen, structural deductions are deductions "that can be described independent of the object language", that is, they are "independent of the language to which the premises and conclusions belong" (see *loc. cit.*). The language would be in this case FOL.

Obviously, Došen's framework is different from the present one and his idea of analysis does not depend on features of the notation; it is essentially analysis of meaning (see Došen's discussion on the subject in [3], 368 ff.). So, two different notions of analysis must be distinguished here. However, Došen's approach can be transposed to Peirce's diagrammatical context, where EGs serve the elucidation (or, better, interpretation) of logical constants of FOL.

Thus, the sign → in the sentence A → B is interpreted by saying that A and B are connected in this formula like they are connected in a scroll, when A is in the outer loop and B is in the inner loop. Furthermore, the sign ¬ in ¬A is interpreted by saying that A plays, in this sentence, the same role that it plays in the scroll when it is in the outer loop and the inner loop is a *pseudograph*. This interpretation of logical constants has been only a sketch because it was limited to the case of some connectives of FOL. It should be extended to quantifiers and the identity sign. In this case, the use of the *line of identity* from the Beta systems of EGs can be unavoidable.

## 5 Conclusions and Final Remarks

From the exploration of the analogies between EGs and Structural Reasoning, it can be conclude that EGs based on the scroll offer a *diagrammatic* way to understand structural deductions and to interpret logical constants, being a *visual* way where specific cognitive skills are implied. This interpretation is based on *a non-analyzable sign*, in the sense of a *diagrammatology*. Here lies the most important conceptual novelty of using the EGs instead of the traditional sequent systems. This interpretation also paves the way to a semiotic analysis (basically based on the well-known Peircean definition of sign in [2] 2.228). For example, it could be argued that the logical constants of FOL designate logical concepts that are visualized through EGs with the scroll. Of course, such claim requires a thorough examination.

Finally, the *illatives graphs* characterized by the scroll seem to be incomplete with respect to classical logic (and it has been suggested that they correspond to intuitionistic logic, see [8] *inter alia*). Presumably, Peirce realizes this fact and ultimately adopted the cut as a basic graph in the system of EGs. A deeper examination of the scroll could lead to the characterization of non-classical logical constants in a similar way to the so-called *substructural* logics.

## References

1. Roberts, D.: The Existential Graphs of Charles. S. Peirce. Mouton, La Haya (1973)
2. Peirce, C.S.: Collected Papers. 8 volumes, vols. 1–6 edited by Hartshorne, C., Weiss, P., vols. 7 and 8 edited by Burks, A.W. Harvard University Press, Cambridge (1931–1958)
3. Došen, K.: Logical constants as punctuation marks. Notre Dame J. Formal Logic **30**(3), 362–381 (1989). https://doi.org/10.1305/ndjfl/1093635154
4. Schroeder-Heister, P.: Resolution and the origins of structural reasoning: early proof-theoretic ideas of Hertz and Gentzen. Bull. Symb. Log. **8**, 246–265 (2002). https://doi.org/10.2178/bsl/1182353872
5. Gentzen, G.: Untersuchungen über das logische Schließen. Mathematische Zeitschrift **39**, 176–210, 405–431 (1935). English Translation in Gentzen, G.: Collected Papers, Transl. and edited By Szabo, M.E., pp. 68–131. North-Holland, Amsterdam-London (1969)
6. Bellucci, F., Pietarinen, A.-V.: Existential graphs as an instrument for logical analysis. Part 1: Alpha. Rev. Symb. Log. **9**(2), 209–237 (2016). https://doi.org/10.1017/S1755020315000362. ISSN 1755-0203
7. Stjernfelt, F.: Two iconicity notions in Peirce's diagrammatology. In: Schärfe, H., Hitzler, P., Øhrstrøm, P. (eds.) ICCS-ConceptStruct 2006. LNCS (LNAI), vol. 4068, pp. 70–86. Springer, Heidelberg (2006). https://doi.org/10.1007/11787181_6
8. Zalamea, F.: Peirce's logic of continuity: existential graphs and non-Cantorian continuum. Rev. Mod. Log. **9**, 115–162 (2003). https://projecteuclid.org/euclid.rml/1081173838

# Multiple Readings of Existential Graphs

Takashi Sasaki(✉)

Kansai University, Osaka, Japan
7sasaki@gmail.com

**Abstract.** In her writings, Shin points out that non-symbolic representation systems have so far been underestimated in favour of symbolic systems. Her alternative aims to overcome the shortcomings of diagrammatic systems while saving the benefits by formalising them in a way that takes advantage of the iconic properties of diagrammatic representations. Specifically, it proposes a natural system by providing a new formulation for reading algorithms and the inference rules of C. S. Peirce's Existential Graphs (EG) that is easy to understand and use. In this study, I cover issues related to multiple readings. In their papers, Bellucci and Pietarinen critically examines Shin's arguments from several points of view. According to Shin, multiple readings are an example that shows the typical characteristics of diagrammatic systems that are not possible in symbolic systems but possible in the alpha part of EG. According to Bellucci and Pietarinen, in contrast, the multiple readings argument is useless to distinguish diagrammatic systems from symbolic ones because it contains circular arguments. Through an examination of this issue, this study considers the diagram and language differences, and differences between the icons and symbols.

**Keywords:** Existential graphs · Multiple readings · C. S. Peirce

## 1 Shin's Argument

C. S. Peirce's Existential Graphs (EG) is analysed by Shin as one of the multi-modal or heterogeneous representation systems that employs both symbolic and diagrammatic elements. As background to her analysis, the following two assumptions are cited ([2], p. 2):

1. Symbolic representation systems can be distinguished from non-symbolic systems, especially diagrammatic ones.
2. Diagrammatic representation systems have thus far been underestimated more than symbolic representation systems.

Shin mainly questioned assumption (2), namely, the prejudice against diagrammatic systems. This underestimation has been brought about by emphasising the disadvantage of diagrammatic systems in spite of their benefits. The typical benefit is the increased efficiency and the disadvantage is the possibility of errors. As the weak points of EG, the difficulty of deciphering it, the user-unfriendly inference rules, and its unnaturalness are cited. Her alternative aims to overcome the shortcomings of diagrammatic systems while saving the benefits by formalising them in a way that takes advantage of the iconic features of EG. Specifically, it proposes a natural system by providing a new

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 598–604, 2018.
https://doi.org/10.1007/978-3-319-91376-6_54

formulation for reading algorithms and the inference rules of EG that is easy to understand and easy to use. In what manner are those iconic features reformulated as benefits of EG? Shin mainly stresses the multiple readings of the Alpha graphs of EG. In this study I consider the points of criticism of the multiple readings argument.

Multiple-Readings Algorithm
Let X and Y be Alpha graphs.

1. If X is an empty space, its translation is $\top$.
2. If X is a sentence letter, its translation is X.
3. If a translation of X is $\alpha$, then a translation of [X] is $\neg\alpha$.
4. If a translation of X is $\alpha$ and a translation of Y is $\beta$, then
    (a) a translation of X Y is $(\alpha \wedge \beta)$,
    (b) a translation of [X Y] is $(\neg\alpha \vee \neg\beta)$,
    (c) a translation of [X [Y]] (i.e., scroll with X in the outside cut and Y in the inner cut) is $(\alpha \rightarrow \beta)$, and
    (d) a translation of [[X] [Y]] is $(\alpha \vee \beta)$ ([2] p. 75).

According to her description, in symbolic systems, multiple readings cause ambiguity. Therefore, we must provide every wffs in the system with a unique reading to avoid ambiguity through syntax and semantics. On the other hand, in EG, we can make multiple readings and it does not cause ambiguity because these readings are logically equivalent to one other.

What causes the difference and ambiguity between symbolic and diagrammatic systems represented by EG in multiple readings? In other words, why are multiple readings possible in EG without causing ambiguity while in symbolic systems a unique readability should be ensured?

Shin says that the linearity is the difference ([3] p. 338). The linearity of the symbolic notations requires unique readability. If the way of separating sequences of various symbols has no restrictions, we cannot avoid ambiguity on the assumption of linearity.

On the other hand, in EG the arrangement of graphs representing facts has no linearity. It is two-dimensional instead of one-dimensional. In this way, multiple readings without ambiguity become possible due to the non-linearity in diagrammatic representation systems. Moreover, according to Shin, multiple readings are not only theoretically feasible in diagrammatic systems, but also enable these systems to be more natural and efficient from a practical point of view.

## 2 Objection to Shin's Argument

Bellucci and Pietarinen insist that there is confusion in this view. According to them, the multiple readings argument could hold under the following two conditions:

Condition 1: 'Reading off' an Alpha graph corresponds to translating it into a formula in ordinary symbolic notation.
Condition 2: In order to generate multiple readings, the target-language must have a richer logical vocabulary than the source-language ([1] p. 21).

That is, from condition (1), the Alpha graph of EG is the source language and the symbolic notation is the target language to which the source language is translated. If condition (2) is not satisfied, it means that the symbolic language as the target language only has two logical connectives of negation and conjunction, like EG as the source language. Multiple readings would then become a unique reading because we would be unable to distinguish alternatives for the reading of the representation in that case.

As to condition (1), Shin refers to the limitations inherent in the approach of translating graphs of EG into language in general ([2] p. 93). However, it seems to be correct at least as a condition for this argument. Bellucci and Pietarinen point out that if the purpose of the argument is to find out the difference between EG and symbolic notation in general, it is problematic.

As can be seen from these conditions, this argument is possible when we take EG with two logical connectives as the source language, and take symbolic language with four connectives as the target language into which EG is translated. Additionally, EG is compared to a symbolic language with four connectives as another source language that is unable to have multiple readings in the target language with four connectives. Finally, whether or not the source languages have such readings is considered to be the difference between both languages. However, considering the purpose of this discussion, EG should be compared to a symbolic language with two of the same connectives of negation and conjunction (2SL). The difference between EG and symbolic notation in general should be found through translating both EG and 2SL into a symbolic language with four connectives. Since we can adjust the algorithm to translate 2SL into 4SL, symbolic language also has multiple readings that meet conditions (1) and (2). Therefore, Bellucci and Pietarinen point out that the multiple readings argument has a third condition:

Condition 3: A formula can have multiple readings in a symbolic target language only if it is not itself symbolic.

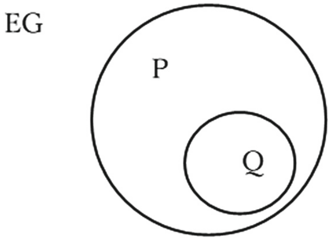
This condition poses a serious difficulty. For Shin's aim is to find a feature that would prove that Alpha is not symbolic, and in general one that distinguishes non-symbolic from symbolic notations. She thinks she has found one in that the former may have multiple readings while the latter cannot. Unfortunately, the notion of multiple readings is here itself defined in terms of such a distinction: when a symbolic language has multiple translations in another, possibly more expressive symbolic language, we do not call these multiple readings. Shin's argument therefore boils down to that Alpha has multiple readings because it is not symbolic, as the multiple-readings algorithm in general only applies when a non-symbolic language is multiply translated into a symbolic language. There is a *circulus in definiendo* here. ([1] p. 23)

This criticism is sharp and convincing. In particular, the idea that a symbolic system has multiple readings is an important suggestion, while Shin seems to suggest that it is only possible in EG. However, do the arguments that focus on multiple readings lose their value by noting this suggestion? Do some characteristics of EG worth noting have no connection with this matter? Through a comparison of Alpha graphs of EG with 2SL, I consider this issue because Bellucci and Pietarinen point out that we should make this comparison.

## 3 Comparison of EG and 2SL with 4SL

In the following example, the graph of the Scroll in EG has three readings in 4SL, which consist respectively of (1) $\neg(P \wedge \neg Q)$, (2) $\neg P \vee Q$, and (3) $P \rightarrow Q$.

As can be seen from this example, according to conditions (1) and (2) there seems to be no problem in understanding formula 2SL as a source language, as well as the Alpha graph of EG (scroll) as having the three available readings in the target language 4SL. To be exact, in the case of 2SL, one of the three in 4SL is exactly the same expression (Fig. 1).



2SL
$\neg(P \wedge \neg Q)$

4SL
(1) $\neg(P \wedge \neg Q)$
(2) $\neg P \vee Q$
(3) $P \rightarrow Q$

**Fig. 1.** Scroll of EG and the readings of the graph in 2SL and 4SL

On this basis, this study asks, can I deduce something from such a comparison? In other words, what can I learn from this comparison about the similarities and differences between 2SL and EG?

What are the similarities between 2SL and EG? Firstly, both notations have two logical connectives, negation and conjunction, and cut and juxtaposition. Secondly, as a result, it will enable multiple readings. As already seen, multiple expressions (1), (2), and (3) in 4SL, which are logically equivalent but do not resemble each other, can be represented in one expression or graph in 2SL and EG.

This could mean that we can grasp the logical equivalence of (1), (2), and (3) without proving it by inference rules if we know how to read these representations. Conversely, since in 4SL different information about (1), (2), and (3) is represented separately by different formulas that are not similar to each other, their equivalence will have to be proven by using inference rules. However, it seems to be difficult to use this feature to make the distinction between diagrammatic and symbolic representations, because the phenomenon itself is available in 2SL as well as EG, as pointed out by Bellucci and Pietarinen.

Thirdly, we can regard a complex expression as one thing that consists of some simple formulas or graphs with two logical connectives shared by 2SL and EG. These complex notations can have the same function as disjunction and conditional connectives, which are the derived connectives in 4SL. As for EG, Shin's multiple readings algorithm 4(d) corresponds to the conjunction which reflects the NNF reading, and 4(c) corresponds to the conditional, which reflects Peirce's 'scroll'.

In this way, this third point is one of the necessary elements needed to realise multiple readings. Presumably, one thing that Shin wants to show is what one can do with EG by using these visually obvious features. According to the previous example, this point can be explained as follows.

Firstly, in the example of the scroll, if we read the Alpha graph from the outside to the inside with regard to every conjunction and negation as a basic element, we can get $\neg(P \wedge \neg Q)$ which corresponds to the formula in 2SL and (1) in 4SL. It is what Peirce calls an 'endoporeutic reading'. Secondly, if we read one complex graph as a disjunction, we find that it is the result of erasing a cut enclosing P from the original graph, and therefore we are able to confirm (2). Thirdly, if we read 'scroll' as conditional, we can confirm (3).

The same can be said about the case of 2SL. If we read $\neg(\neg P \wedge \neg Q)$ as a disjunction, $\neg(P \wedge \neg Q)$ as conditional, we can confirm (2) and (3) in 4SL. Following Norman, we call this a molecular reading [4].

It is clear that the third property has a relation with multiple readings. Both in EG and 2SL we can grasp the logical equivalence of (1), (2), and (3) without proving it according to inference rules if we know how to read these representations. From this point we can also see the merit of the four connectives in 4SL. The information of (2) and (3) can be expressed without using brackets in 4SL. This is the important difference between 2SL and 4SL.

In these cases, brackets are necessary when ambiguity concerning the scope of logical connectives may occur without it. In the case of the conditional connective, if we removed the brackets from formula 2SL, the scope of the negation of the original formula is different from the original '$\neg(P \wedge \neg Q)$'. However, in 4SL we can remove the brackets by adding the derived connectives of disjunction and conditional. Although this seems to be trivial at first, it is related to important issues as will be described later.

## 4 Comparison of EG with 2SL

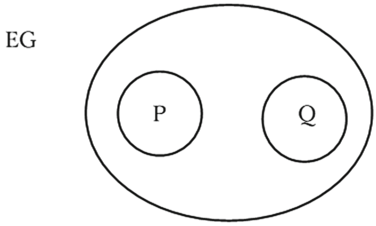
What are the differences between 2SL and EG? Firstly, if we consider the propositional symbol as a starting point, there are differences in the patterns of the representation that are required within propositional logic. In the case of EG, there are only two patterns, that is, enclosing the propositional symbol or graph by a cut and juxtaposing propositional symbols. In contrast, 2SL has the negation symbol ($\neg$) and brackets that correspond to the cut of EG, and the conjunction symbol ($\vee$) and brackets that correspond to the juxtaposition of EG.

In this way, there are differences between both concerning the apparatus and the way in which it realises logical representation. In particular, in order to make a new proposition from any proposition, EG requires only the cut, but 2SL requires a negation symbol, conjunction symbol, and brackets. In other words, in order to determine what new propositions can be made from arbitrary propositions, the following discernible differences arise with the exception of the existence of the propositional symbol.

That is, in the case of 2SL, it is necessary to discriminate whether a negative sign is attached, a conjunctive sign is attached, and whether brackets are present, and if so, which range is being sandwiched. In contrast, in the case of EG, all we need to discern is whether or not it is surrounded by cuts.

This difference causes the differences of the object for molecular reading in 2SL and EG. The objects for molecular reading are complex formulas or graphs, but in EG the objects form a shape consisting of just two elements of propositional symbols and cuts. In 2SL, they are complex formulas linear in nature, consisting of the four elements listed above.

In particular, a composite graph as a whole is grouped by the two-dimensional enclosure of a cut in EG, which makes it easier to understand, write, and memorise than representation using brackets to specify scopes. In general, a difference arises in the ease of use. If we compare it to each of the following, in the practice of molecular reading, there seems to be differences in these regards between 2SL and EG. This difference seems to be related to the naturalness and efficiency seen from the practical point of view mentioned earlier (Fig. 2).



2SL
$\neg(\neg P \wedge \neg Q)$

4SL
(1) $\neg(\neg P \wedge \neg Q)$
(2) $P \vee Q$
(3) $\neg P \rightarrow Q$
(4) $\neg Q \rightarrow P$

**Fig. 2.** The graph of EG corresponding to disjunction and the readings in 2SL and 4SL

Secondly, this relates to the first point, 2SL uses brackets but EG does not need it. The cut of EG plays the role of the logical connective of negation that negates the figure surrounded by it, and at the same time determines the scope of the negation.

However, in the case of 2SL as described above, it is necessary to determine the scope of the negative symbol using brackets if necessary. This not only means that EG requires less necessary apparatus for representation than 2SL, but also that there is an important difference between 2SL and EG. That is, the boundaries of the scope indicated by the cut will be more easily distinguished visually than that indicated by brackets. This difference comes from two facts. One is that the brackets are aligned linearly as well as in terms of character. The other is that they must make a pair in order to indicate the scope. The larger the number of brackets, the more prominent this difference becomes.

It was due to the merit of adding derived connectives in 4SL that logical information is expressed as much as possible without the use of brackets. In the case of 4SL, in exchange for this benefit it causes the necessity of using inference rules to prove equivalence. However, we can avoid this disadvantage in EG through multiple readings. Therefore, in that regard, we can take advantage of the 4SL without losing the advantages of the 2SL.

## 5 Conclusion

Thus far I have recognised 2SL as source languages as well as EG, and found similarities between 2SL and EG by comparing them to 4SL. I then examined the differences between 2SL and EG.

Through the discussion of multiple readings, Shin tries to state the difference between EG as being representative of the graphic system and 4SL as a symbolic system. Certainly, according to this discussion, it cannot be argued that EG and 2SL are also different based on the possibility of multiple readings. In this sense, it can be said that Bellucci and Pietarinen's criticism clarifies why multiple readings are possible.

However, by comparing EG and 2SL, several differences were clarified as to how to realise multiple readings. Firstly, there is a difference regarding tool construction and how to realise logical representation. Secondly, the necessity of using brackets is another difference. These differences are features of EG, unlike the general symbolic system, and these features are considered to be related to practical advantages. Therefore, the discussion of multiple readings can play a role in highlighting such features of EG.

## References

1. Bellucci, F., Pietarinen, A.-V.: Existential graphs as an instrument of logical analysis: part 1. Alpha. Rev. Symb. Log. **9**(2), 209–237 (2016)
2. Shin, S.-J.: The Iconic Logic of Peirce's Graphs. The MIT Press, Cambridge (2002)
3. Shin, S.-J.: Peirce's alpha graphs and propositional language. Semiotica **186**(1/4), 333–346 (2011)
4. Norman, J.: Diagrammatic Reasoning and Propositional Logic. University College London, London (1999)

# Logic and Diagrams

# Towards a Proof Theory for Heterogeneous Logic Combining Sentences and Diagrams

Ryo Takemura(✉)

Nihon University, Tokyo, Japan
takemura.ryo@nihon-u.ac.jp

**Abstract.** We attempt to develop a proof theory for heterogeneous logic combining first-order formulas and diagrams. In proof theory, normal proofs and normalization play a central role, which makes it possible to analyze and characterize the structure of proofs in a given system. In light of the difference between linguistic reasoning and diagrammatic reasoning, we apply the traditional proof theory developed in symbolic logic to heterogeneous logic, and we give a characterization of the structure of heterogeneous proofs based on our normalization theorem.

## 1 Introduction

Heterogeneous reasoning combining various graphical/diagrammatic and sentential/linguistic representations has been an important subject in the study of diagrammatic reasoning, and several heterogeneous systems have been investigated so far. Blocks world systems [1,3]; Euler and Venn systems [6,14,17]; and correspondence table systems [2,16] are some examples of such studies. However, proof theory of heterogeneous logic has not yet been developed much.

Existing development in proof theory has taken place by investigating logical proofs based on sentential/linguistic representation. One of the major goals of proof theory is to analyze and characterize the structure of proofs, and thereby investigate effective strategies to construct/search proofs in a system. Thus, proof theory offers a basis in theorem proving. In such proof theory, normal proofs and normalization play a central role. Thanks to the normalization theorem, any proof is reduced to a normal form, and we are able to focus on normal proofs for their analysis and characterization. Therefore, Gentzen [5] called the normalization theorem as Hauptsatz (Main Theorem) of proof theory.

If we translate diagrams into formulas of first-order logic (FOL), we can apply the usual proof-theoretic techniques to heterogeneous/diagrammatic logic in a straightforward manner. Based on this idea, the author in [15] investigated a class of Euler diagrammatic proofs called "N-normal diagrammatic proofs" that has a one-to-one correspondence with normal proofs in FOL. Although N-normal diagrammatic proofs have the structure of linguistic FOL proofs, they do not

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 607–623, 2018.
https://doi.org/10.1007/978-3-319-91376-6_55

reflect characteristics of diagrammatic proofs. Thus, by using such an N-normal form, it is difficult to characterize the structure of diagrammatic proofs in a general sense.

There is a major difference between linguistic reasoning and diagrammatic reasoning with respect to their methods and strategies. Linguistic reasoning, as characterized by the normalization theorem (cf. [10,11]) for FOL, consists of (1) decomposition of given premises, and (2) construction of a conclusion by combining the decomposed formulas. In contrast, diagrammatic reasoning consists of (1) construction of a (maximal) diagram by unifying pieces of information contained in given premises, and (2) extraction of a conclusion from the unified diagram (see, for example, [13]). In light of such a distinction, we apply and extend the traditional proof theory developed in symbolic logic, and we give a characterization of the structure of heterogeneous proofs based on our normalization theorem. We study heterogeneous logic combining first-order formulas and diagrams within the framework of natural deduction. We investigate abstract properties of heterogeneous proofs independent of particular systems. Such properties are shared by various concrete systems such as Euler and Venn systems (e.g., [6,7,9,12]); blocks world systems [1,3]; and correspondence table systems (e.g., [2,16]). In Sect. 2, we describe our abstract syntax of heterogeneous logic. In Sect. 3, we introduce our inference rules. Rules for formulas are the usual natural deduction rules for FOL. We investigate, among various inference rules, heterogeneous rules Apply and Observe (cf. [1,3,6]), as well as diagrammatic rules Unification and Deletion (cf. [7,9,12]) exclusively, since these rules are considered to be the most basic rules and are shared by various heterogeneous systems. In Sect. 4, we investigate a normalization theorem in our heterogeneous system, and provide a characterization for the structure of our heterogeneous proofs.

## 2 Syntax of Heterogeneous Logic

We introduce syntax of heterogeneous logic abstractly. While concrete syntax is defined in each system, here we extract common items to be specified in each system. The syntax of heterogeneous logic is defined by specifying the following *formulas, diagrams, diagrammatic objects*, and *diagrammatic formulas*:

**Formulas:** denoted by $\varphi, \psi, \sigma, \varphi_1, \varphi_2, \ldots$. Formulas of FOL (first-order logic) are defined inductively as usual:

$$\varphi ::= A(t_1, \ldots, t_n) \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \varphi \to \varphi \mid \neg\varphi \mid \bot \mid \forall x \varphi \mid \exists x \varphi$$

where $A(t_1, \ldots, t_n)$ is an atomic formula consisting of a predicate $A$ and terms $t_1, \ldots, t_n$. When $A$ is a unary predicate, we usually omit parentheses and write an atomic formula such as $At$.

**Diagrams:** denoted by $\mathcal{D}, \mathcal{E}, \mathcal{F}, \mathcal{D}_1, \mathcal{D}_2, \ldots$. The answer to what qualifies as a concrete diagram depends on each system. Cf. Example 1 below.

**Diagrammatic Objects:** denoted by $o, o_1, o_2, \ldots$. They are components of diagrams, and the answer to what qualifies as a concrete diagrammatic object depends on each system. By $ob(\mathcal{D})$, we denote the set of diagrammatic objects that appear on a diagram $\mathcal{D}$.

For example, diagrammatic objects are named circles and points, linking between points, shading in Euler and Venn systems; and blocks such as cubes and tetrahedron in blocks world systems.

**Diagrammatic Formulas:** denoted by $\varphi^d, \psi^d, \sigma^d, \varphi^d_1, \varphi^d_2 \dots$. They describe pieces of basic information contained in diagrams. The answer to what kind of formulas qualify as diagrammatic formulas is specified in each system.

For example, in a Venn diagrammatic system, $\neg\exists x(A_1x \wedge \cdots \wedge A_nx \wedge \neg B_1x \wedge \cdots \wedge \neg B_mx)$ is a diagrammatic formula describing that the region inside circles $A_1, \dots, A_n$ and outside $B_1, \dots, B_m$ is shaded. See Example 1 for an example of an Euler system. In a blocks world system, atomic formulas of hyperproof such as $small(t), cube(t), leftof(t_1, t_2)$, and so on are diagrammatic formulas.

We identify a diagrammatic object in a diagram with a predicate or a term describing the object in a diagrammatic formula. Thus, for a diagrammatic formula $\varphi^d$, we use $ob(\varphi^d)$ to denote the set of predicates and terms appearing in $\varphi^d$, each of which expresses a diagrammatic object. We further identify a relation holding on a diagram with a diagrammatic formula that describes the relation.

For a diagram $\mathcal{D}$, $type(\mathcal{D})$ is the set $\{\varphi^d_1, \varphi^d_2, \dots, \varphi^d_n\}$ of diagrammatic formulas, such that the relation $\varphi^d_i$ holds on $\mathcal{D}$ if and only if $\varphi^d_i \in type(\mathcal{D})$. We identify $type(\mathcal{D})$ with the conjunctive formula $\varphi^d_1 \wedge \varphi^d_2 \wedge \cdots \wedge \varphi^d_n$. The *type* of a diagram is the symbolic specification of the diagram.

Based on the above specification, we elucidate our **postulates** in this article.

1. We presume the set of diagrammatic formulas to be a subset of FOL formulas. Thus, every piece of basic information contained in the diagrams is described by a formula of FOL.
2. We regard a diagram, in view of FOL, as the conjunction of diagrammatic formulas comprising the diagram. Thus, we do not consider linking between diagrams in this article, since linking between diagrams makes its type disjunctive. (Although we allow linking between points.)
3. We presume $type(\mathcal{D})$ is *deductively closed* with respect to diagrammatic formulas. That is, when $type(\mathcal{D}) = \{\varphi^d_1, \varphi^d_2, \dots, \varphi^d_n\}$, if $\varphi^d_1 \wedge \varphi^d_2 \wedge \cdots \wedge \varphi^d_n$ implies a diagrammatic formula $\psi^d$ such that $ob(\psi^d) \subseteq ob(\mathcal{D})$, then $\psi^d \in type(\mathcal{D})$. The above "implies" is considered as an appropriate semantic consequence or syntactic consequence in FOL.

Although we do not enter into detail, the semantics of our heterogeneous system is defined as the usual set-theoretic semantics for FOL, since our diagram corresponds to a conjunction of diagrammatic formulas.

Although we illustrate only one concrete Euler diagrammatic system of [9] below because of space limitation, our definition is valid for other Euler and Venn systems (e.g., [6,7,12]); blocks world systems [1,3]; correspondence table systems (e.g., [2,16]), and so on.

*Example 1 (*EUL*-diagrams).* An Euler diagram of [9], called an EUL-diagram, is defined as a plane with named circles and points. Each EUL-diagram is specified by inclusion and exclusion relations maintained between circles and points on the diagram. EUL-diagrams can express neither disjunctive information with respect to the location of a point, nor information of contradiction.

**Diagrams.** An EUL-diagram is a plane with a finite number of (named) simple closed curves (simply called (named) circles and denoted by $A, B, C, \dots$), constant points (denoted by $a, b, c, \dots$), and existential points (denoted by $x, y, z, \dots$). Constant points and existential points are collectively called (named) points, and are denoted by $t, s, t_1, t_2, \dots$.

**Diagrammatic objects** are named circles and points.

**Diagrammatic formulas.** An EUL-diagram is specified in terms of topological relations $\sqsubset$ (*inside of*), $\vdash\!\dashv$ (*outside of*), and $\bowtie$ (*crossing*) between diagrammatic objects holding on the diagram. These relations are expressed by the following diagrammatic formulas:

- $\forall x(Ax \to Bx)$ for $A \sqsubset B$ ($A$ is inside of $B$);
- $\forall x(Ax \to \neg Bx)$ for $A \vdash\!\dashv B$ ($A$ is outside of $B$);
- $\forall x(Ax \to Ax) \wedge \forall x(Bx \to Bx)$ for $A \bowtie B$ (there is at least one crossing point between $A$ and $B$);
- $At$ for $t \sqsubset A$ ($t$ is inside of $A$);
- $\neg At$ for $t \vdash\!\dashv A$ ($t$ is outside of $A$).

For example, the topmost diagram in Fig. 1 of Example 2 consists of EUL-relations $A \sqsubset B, A \vdash\!\dashv E, B \bowtie E$, and hence, its type is $\{\forall x(Ax \to Bx), \forall x(Ax \to \neg Ex), \forall x(Bx \to Bx) \wedge \forall x(Ex \to Ex)\}$.

## 3 Inference Rules of Heterogeneous Logic

We first review the usual inference rules of natural deduction for FOL in Sect. 3.1. Then, in Sect. 3.2, we introduce our heterogeneous inference rules Apply and Observe, as well as purely diagrammatic inference rules Unification and Deletion, which are shared in the typical heterogeneous systems. In contrast to a linguistic FOL rule, whose conclusion is always well-defined given well-defined premises, a diagrammatic rule's conclusion may not be defined even if the premises are well-defined, because of the expressive limitations of diagrams.

### 3.1 Natural Deduction Rules for FOL

A proof in natural deduction is structured as a tree consisting of formulas as its nodes and the following inference rules as its edges. The top formulas of the tree are the assumptions, and the other formulas of the tree follow from the formulas immediately above, using one of the rules. A formula $A$ in the tree is said to depend on the assumptions standing above $A$ that have not been closed by some inference preceding $A$. In the following rules, a formula written within square brackets indicates that, the assumptions of this form occurring above the premises are closed at the inference. See [4,10,11] for a detailed introduction to natural deduction.

**Definition 1 (Rules for FOL).** The natural deduction rules for FOL consist of the following dual pairs, each pair consisting of an *introduction (I)* and an *elimination (E)* rules, for each connective $\wedge, \vee, \rightarrow, \neg, \forall, \exists$, as well as $\bot E$ and $RAA$:

$$\dfrac{\begin{matrix}\vdots & \vdots \\ \varphi & \psi\end{matrix}}{\varphi \wedge \psi}\wedge I \qquad \dfrac{\begin{matrix}\vdots \\ \varphi_1 \wedge \varphi_2\end{matrix}}{\varphi_i}\wedge E_{(i\,=\,1,\,2)} \qquad \dfrac{\begin{matrix}\vdots \\ \varphi_i\end{matrix}}{\varphi_1 \vee \varphi_2}\vee I_{(i\,=\,1,\,2)} \qquad \dfrac{\begin{matrix} & [\varphi]^n & [\psi]^n \\ \vdots & \vdots & \vdots \\ \varphi \vee \psi & \sigma & \sigma\end{matrix}}{\sigma}\vee E, n$$

$$\dfrac{\begin{matrix}[\varphi]^n \\ \vdots \\ \psi\end{matrix}}{\varphi \rightarrow \psi} \rightarrow I, n \qquad \dfrac{\begin{matrix}\vdots & \vdots \\ \varphi & \varphi \rightarrow \psi\end{matrix}}{\psi} \rightarrow E \qquad \dfrac{\begin{matrix}[\varphi]^n \\ \vdots \\ \bot\end{matrix}}{\neg\varphi}\neg I, n \qquad \dfrac{\begin{matrix}\vdots & \vdots \\ \varphi & \neg\varphi\end{matrix}}{\bot}\neg E$$

$$\dfrac{\begin{matrix}\vdots \\ \varphi(x)\end{matrix}}{\forall x \varphi(x)}\forall I \quad \dfrac{\begin{matrix}\vdots \\ \forall x \varphi(x)\end{matrix}}{\varphi(t)}\forall E \quad \dfrac{\begin{matrix}\vdots \\ \varphi(t)\end{matrix}}{\exists x \varphi(x)}\exists I \quad \dfrac{\begin{matrix} & [\varphi(x)]^n \\ \vdots & \vdots \\ \exists x \varphi(x) & \psi\end{matrix}}{\psi}\exists E, n \quad \dfrac{\begin{matrix}\vdots \\ \bot\end{matrix}}{\varphi}\bot E \quad \dfrac{\begin{matrix}[\neg\varphi]^n \\ \vdots \\ \bot\end{matrix}}{\varphi}RAA, n$$

In $\forall I$, the variable $x$ may not occur freely in any open assumption, on which $\varphi(x)$ depends; in $\exists E$, $x$ may not occur freely in $\psi$ nor in any open assumption on which $\psi$ depends, except in $\varphi(x)$.

### 3.2 Heterogeneous Rules

As representative rules of heterogeneous systems independent of specific diagrams, we investigate the following rules, where app and obs consist of the dual pair of heterogeneous rules, and uni and del consist of the dual diagrammatic rules.

**Definition 2.** Heterogeneous rules of Apply (app) and Observe (obs), and diagrammatic rules of Unification (uni) and Deletion (del) have the following forms:

$$\dfrac{\begin{matrix}\vdots & \vdots \\ \mathcal{D} & \varphi^d\end{matrix}}{\mathcal{D} + \varphi^d}\,\mathsf{app} \qquad \dfrac{\begin{matrix}\vdots \\ \mathcal{D}\end{matrix}}{\psi^d}\,\mathsf{obs} \qquad \dfrac{\begin{matrix}\vdots & \vdots \\ \mathcal{D} & \mathcal{E}\end{matrix}}{\mathcal{D} + \mathcal{E}}\,\mathsf{uni} \qquad \dfrac{\begin{matrix}\vdots \\ \mathcal{D}\end{matrix}}{\mathcal{D} - \{o_1, \ldots, o_n\}}\,\mathsf{del}$$

where $\psi^d \in type(\mathcal{D})$ in obs, and $o_1, \ldots, o_n \in ob(\mathcal{D})$ in del.
These rules are applicable when every $\mathcal{D} + \varphi^d, \mathcal{D} + \mathcal{E}, \mathcal{D} - \{o_1, \ldots, o_n\}$ is a well-defined diagram.

In del, $\mathcal{D} - \{o_1, \ldots, o_n\}$ is the diagram obtained by deleting diagrammatic objects $o_1, \ldots, o_n$ from $\mathcal{D}$. In app, $\mathcal{D} + \varphi^d$ is the diagram that extends from $\mathcal{D}$ by adding the information of $\varphi^d$ (cf. [1,3]). In uni, $\mathcal{D} + \mathcal{E}$ is the unified diagram of $\mathcal{D}$ and $\mathcal{E}$ (cf. [7,9]). Depending on the specific definition of diagrams in each system, $\mathcal{D} + \varphi^d$ and $\mathcal{D} + \mathcal{E}$ are not always defined. There may be various constraints on app and uni in order to avoid the case where its conclusion is undefined. Two of the major constraints are that for *indeterminacy* and for *contradiction* as seen in [9]. In this article, we presume that app as well as uni are applicable when $\mathcal{D} + \varphi^d$ (resp. $\mathcal{D} + \mathcal{E}$) is defined as a single diagram. This allows us to exclude the case where several distinguishable diagrams or linking of them is needed (as [7,12]) for representing $\mathcal{D} + \varphi^d$ (resp. $\mathcal{D} + \mathcal{E}$). We also do not take the rule of Cases Exhaustive [1,3] into consideration in this article.

Applications of our inference rules are illustrated in Fig. 1 in Example 2.

A **heterogeneous proof**, denoted by $\pi, \pi_1, \pi_2, \ldots$, is defined inductively as a tree consisting of formulas and diagrams as its nodes, and inference rules as its edges. We write $\alpha_1, \ldots, \alpha_n \vdash \alpha$, when $\alpha$ is provable from premises $\alpha_1, \ldots, \alpha_n$, where $\alpha_i$ is a formula or a diagram.

## 4 Normalization of Heterogeneous Proofs

We review the notions of *detour*, *reduction*, and *normal proof* in the usual natural deduction for FOL (cf. [10,11]) in Sect. 4.1. Then, we discuss their counterparts for our heterogeneous system in Sect. 4.2. In Sect. 4.3, we prove our normalization theorem of heterogeneous proofs. Based on the theorem, we investigate a characterization of the structure of heterogeneous proofs in Sect. 4.4.

### 4.1 Normal Proof and Normalization in FOL

In general, a natural deduction proof may contain some redundant steps and formulas called **maximal formulas**, i.e., formulas that stands at the same time as the conclusion of an introduction rule and as the major premise of an elimination rule. For example, the formula $\varphi_1 \wedge \varphi_2$ and the pair of applications of $\wedge I$ and $\wedge E$ rules on the left in the following proof are redundant, because without them we already have a proof $\pi_1$ of $\varphi_1$ as illustrated on the right.

$$\dfrac{\dfrac{\begin{array}{c}\vdots\ \pi_1\\ \varphi_1\end{array}\qquad\begin{array}{c}\vdots\ \pi_2\\ \varphi_2\end{array}}{\varphi_1 \wedge \varphi_2}\,\wedge I}{\varphi_1}\,\wedge E \qquad\qquad\qquad \begin{array}{c}\vdots\ \pi_1\\ \varphi_1\end{array}$$

A maximal formula along with its related pair of applications of an introduction and an elimination rule are together called **detour** in a proof, and it is possible to remove such a detour as illustrated above. This rule of rewriting a given proof by removing a detour is called the **reduction rule**, and it is defined for every pair of the dual introduction and elimination rules. In addition to the above

$\wedge$-reduction rule, the reduction rules for $\to, \vee, \forall, \exists$ are defined as follows, where the part in a given proof on the left is rewritten into a form on the right:

- $\to$-reduction

$$\dfrac{\dfrac{\begin{matrix}[\varphi]\\ \vdots\ \pi_1\\ \psi\end{matrix}}{\varphi\to\psi}\to I \qquad \begin{matrix}\vdots\ \pi_2\\ \varphi\end{matrix}}{\psi}\to E \quad \rhd \quad \begin{matrix}\vdots\ \pi_2\\ \varphi\\ \vdots\ \pi_1\\ \psi\end{matrix}$$

- $\vee$-reduction

$$\dfrac{\dfrac{\begin{matrix}\vdots\ \pi\\ \varphi_i\end{matrix}}{\varphi_1\vee\varphi_2}\vee I \qquad \begin{matrix}[\varphi_1]\\ \vdots\ \pi_1\\ \sigma\end{matrix} \qquad \begin{matrix}[\varphi_2]\\ \vdots\ \pi_2\\ \sigma\end{matrix}}{\sigma}\vee E \quad \rhd \quad \begin{matrix}\vdots\ \pi\\ \varphi_i\\ \vdots\ \pi_i\\ \sigma\end{matrix} \quad (i=1,2)$$

- $\forall$-reduction

$$\dfrac{\dfrac{\begin{matrix}\vdots\ \pi(x)\\ \varphi(x)\end{matrix}}{\forall x\varphi(x)}\forall I}{\varphi(t)}\forall E \quad \rhd \quad \begin{matrix}\vdots\ \pi(t)\\ \varphi(t)\end{matrix}$$

- $\exists$-reduction

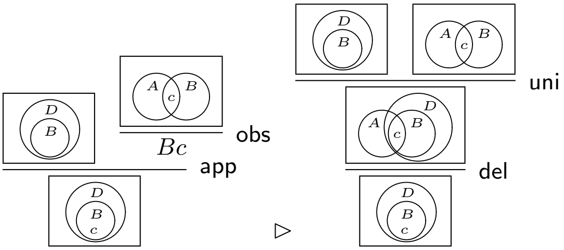
$$\dfrac{\dfrac{\begin{matrix}\vdots\ \pi_1\\ \varphi(t)\end{matrix}}{\exists x\varphi(x)}\exists I \qquad \begin{matrix}[\varphi(x)]\\ \vdots\ \pi_2(x)\\ \sigma\end{matrix}}{\sigma}\exists E \quad \rhd \quad \begin{matrix}\vdots\ \pi_1\\ \varphi(t)\\ \vdots\ \pi_2(t)\\ \sigma\end{matrix}$$

These are the main reduction rules, and see [4,11] for other technical rules. A natural deduction proof is said to be in **normal form** when it does not contain any redex, i.e., detour. Then, the following **normalization theorem** holds: *If $\varphi$ is provable from a set of formulas $\Gamma$, then there is a normal proof of $\varphi$ from $\Gamma$.*

Normalization theorem makes various proof-theoretic analyses possible. For example, the notion of normal proofs enables us to characterize the structure of proofs in a formal system. Prawitz [10,11] shows that each normal proof consists of two parts: (1) an *analytical part* in which premises are decomposed into their components by using elimination rules; (2) a *synthetic part* in which the final components obtained in the analytical part are put together for constructing the conclusion using introduction rules.

## 4.2 Reduction Rules for Heterogeneous Proofs

The notion of reduction in natural deduction for FOL is explained as the removal of a *detour*, i.e., a redundant maximal concept (formula) as well as its introduction and elimination rules. Let us consider what a detour inherent in our diagrammatic inference is. Diagrammatic inference can be characterized by constructing a (maximal) diagram and extracting a conclusion from the diagram. (See, for example, [13].) In such a diagrammatic inference, the use of redundant subconcepts (diagrams/formulas) may be considered as a detour. This detour is part of a proof where one infers by deducing subconcepts, even though one can infer directly by using a superior concept. For example, $Bc$ and the pair of obs-app in the following proof on the left are redundant, i.e., a detour, since the information of $Bc$ is already contained in the diagram above the obs, and we can obtain the same conclusion by directly unifying two premise diagrams without deducing the formula $Bc$.



$$\dfrac{\begin{array}{cc} & \vdots \\ & \dfrac{\mathcal{D}}{\varphi}\,\mathsf{obs} \\ \vdots & \vdots \\ \mathcal{E} & \psi \end{array}}{\mathcal{E}+\psi}\,\mathsf{app} \quad \rhd \quad \dfrac{\dfrac{\begin{array}{cc}\vdots & \vdots\\ \mathcal{E} & \mathcal{D}\end{array}}{\mathcal{E}+\mathcal{D}}\,\mathsf{uni}}{\mathcal{E}+\psi}\,\mathsf{del}$$

This kind of a detour and its reduction can be formulated by the above rule. When $\psi$ depends only on $\varphi$ (i.e., $\varphi \vdash \psi$), $\varphi$ and $\psi$ as well as formulas/diagrams between them are redundant subconcepts. This is because our diagrams are deductively closed, and we are able to infer $\mathcal{E}+\psi$ by directly unifying $\mathcal{E}$ and $\mathcal{D}$ without deducing $\varphi$ nor $\psi$ as illustrated on the right. Note that this reduction is possible under the following conditions: (1) $\mathcal{E}+\mathcal{D}$ is defined as a legal diagram; (2) $\psi$ depends only on $\varphi$ ($\varphi \vdash \psi$). In other words, when we focus and cut out the part from $\varphi$ to $\psi$, it stands as a legal proof independent of the other part of the given proof. Otherwise, we cannot deduce $\psi$ from $\mathcal{D}$, as well as $\mathcal{E}+\psi$ from $\mathcal{E}+\mathcal{D}$ after the reduction.

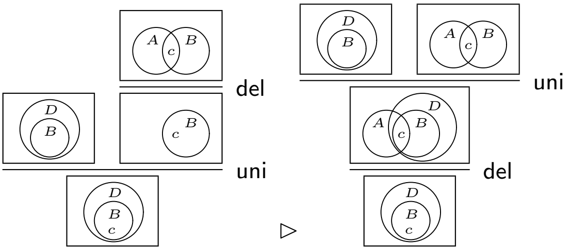
Note that the notion of "detour" is conceptual, and it is not necessarily related to the length of proofs. It is known, in symbolic logic proof theory, that normal proofs may be more lengthy and complex than non-normal proofs, but they are conceptually simpler in the sense that no detour is contained.

The above obs-app-reduction is generalized as follows:
When (1) $\mathcal{E}+\sum\left(\sum\mathcal{D}_n+\sigma_l^d\right)$ is defined, where $\sum\mathcal{D}_n$ is the unified diagram $\mathcal{D}_1+\mathcal{D}_2+\cdots+\mathcal{D}_n$; and (2) $\varphi_1,\ldots,\varphi_n,\sigma_1^d,\ldots,\sigma_l^d \vdash \psi$, the following part in a proof on the left is reduced to the part on the right:

$$\dfrac{\begin{array}{ccc} & \begin{array}{c}\vdots\\ \dfrac{\mathcal{D}_1}{\varphi_1}\end{array}\mathsf{obs}\cdots\begin{array}{c}\vdots\\ \dfrac{\mathcal{D}_n}{\varphi_n}\end{array}\mathsf{obs} & \begin{array}{cc}\vdots&\vdots\\ \sigma_1^d\cdots&\sigma_l^d\end{array}\\ \vdots & \vdots & \\ \mathcal{E} & \psi & \end{array}}{\mathcal{E}+\psi}\,\mathsf{app} \quad \rhd \quad \dfrac{\dfrac{\begin{array}{cc}\vdots & \\ \mathcal{E} & \dfrac{\dfrac{\begin{array}{cc}\vdots&\vdots\\ \mathcal{D}_1\cdots&\mathcal{D}_n\end{array}}{\sum\mathcal{D}_n}\,\mathsf{uni} \quad \begin{array}{cc}\vdots&\vdots\\ \sigma_1^d\cdots&\sigma_l^d\end{array}}{\sum\left(\sum\mathcal{D}_n+\sigma_l^d\right)}\,\mathsf{app}\end{array}}{\mathcal{E}+\sum\left(\sum\mathcal{D}_n+\sigma_l^d\right)}\,\mathsf{uni}}{\mathcal{E}+\psi}\,\mathsf{del}$$

In the above reduced proof on the right, the double line ═══ uni means repeated applications of uni to $\mathcal{D}_1,\ldots,\mathcal{D}_n$; similarly for ═══ app. Each $\sigma_i^d$ is a diagrammatic formula independent of obs. Since $\psi$ may depend not only on diagrams $\mathcal{D}_1,\ldots,\mathcal{D}_n$ but also on formulas $\sigma_1^d,\ldots,\sigma_l^d$, the similar structure of our detour may occur in such a part. Thus, we generalize our reduction by including $\sigma_i^d$ to reduce such a part; cf. Fig. 1 of Example 2.

Let us further consider the dual pair del and uni of diagrammatic rules. The following proof on the left consisting only of diagrams, is considered to contain a similar detour as the obs-app pair. Without deducing the diagram of $Bc$ by deleting circle $A$, we may directly unify the given two premise diagrams as illustrated on the right.



$$\dfrac{\begin{matrix}\vdots\\ \mathcal{E}\end{matrix} \quad \begin{matrix}\vdots\\ \dfrac{\mathcal{F}}{\mathcal{F}'}\,\mathsf{del}\\ \vdots\\ \mathcal{D}\end{matrix}}{\mathcal{E}+\mathcal{D}}\,\mathsf{uni} \quad \rhd \quad \dfrac{\dfrac{\begin{matrix}\vdots\\ \mathcal{E}\end{matrix}\quad\begin{matrix}\vdots\\ \mathcal{F}\end{matrix}}{\mathcal{E}+\mathcal{F}}\,\mathsf{uni}}{\mathcal{E}+\mathcal{D}}\,\mathsf{del}$$

Thus, we can also formulate a reduction rule for a del-uni pair, where, as in the obs-app-reduction, we assume that (1) $\mathcal{E}+\mathcal{F}$ is defined; and (2) $\mathcal{D}$ depends only on $\mathcal{F}'$, i.e., $\mathcal{F}' \vdash \mathcal{D}$.

Since del and obs share a similar structure (i.e., extraction of information), a pair of del and app may form the same detour as before, although del and app are not in duality. Thus, for removing all of the detours of a same kind in a proof, we generalize our reduction to obs/del-app-reduction and del/obs-uni-reduction as follows.

**Definition 3.** obs/del-app-reduction and del/obs-uni-reduction are defined as follows.

- obs/del-app-**reduction (1)**
When $\overrightarrow{\varphi_n}, \overrightarrow{\mathcal{F}'_m}, \overrightarrow{\sigma_l^d} \vdash \psi$ and $\mathcal{E}+\sum\left(\sum\left(\sum D_n+F_m\right)+\sigma_l^d\right)$ is defined in the following part in a proof:

$$\dfrac{\begin{matrix}\vdots\ \pi'\\ \mathcal{E}\end{matrix} \qquad \begin{matrix} \dfrac{\begin{matrix}\vdots\ \pi_1^1\\ \mathcal{D}_1\end{matrix}}{\varphi_1}\,\mathsf{obs}\cdots\dfrac{\begin{matrix}\vdots\ \pi_n^1\\ \mathcal{D}_n\end{matrix}}{\varphi_n}\,\mathsf{obs}\cdots\dfrac{\begin{matrix}\vdots\ \pi_1^2\\ \mathcal{F}_1\end{matrix}}{\mathcal{F}_1'}\,\mathsf{del}\cdots\dfrac{\begin{matrix}\vdots\ \pi_m^2\\ \mathcal{F}_m\end{matrix}}{\mathcal{F}_m'}\,\mathsf{del}\cdots\begin{matrix}\vdots\ \pi_1^3\\ \sigma_1^d\end{matrix}\quad\cdots\begin{matrix}\vdots\ \pi_l^3\\ \sigma_l^d\end{matrix}\\ \vdots\ \pi''\\ \psi\end{matrix}}{\mathcal{E}+\psi}\,\mathsf{app}$$

where $n \neq 0$ or $m \neq 0$, it is reduced to:

$$\dfrac{\dfrac{\begin{matrix}\vdots\ \pi'\\ \mathcal{E}\end{matrix} \qquad \dfrac{\dfrac{\dfrac{\begin{matrix}\vdots\ \pi_1^1\\ \mathcal{D}_1\end{matrix}\quad\cdots\begin{matrix}\vdots\ \pi_n^1\\ \mathcal{D}_n\end{matrix}}{\overline{\sum\mathcal{D}_n}}\,\mathsf{uni}\qquad\begin{matrix}\vdots\ \pi_1^2\\ \mathcal{F}_1\end{matrix}\quad\cdots\begin{matrix}\vdots\ \pi_m^2\\ \mathcal{F}_m\end{matrix}}{\overline{\sum\left(\sum\mathcal{D}_n+\mathcal{F}_m\right)}}\,\mathsf{uni}\qquad\begin{matrix}\vdots\ \pi_1^3\\ \sigma_1^d\end{matrix}\quad\cdots\begin{matrix}\vdots\ \pi_l^3\\ \sigma_l^d\end{matrix}}{\overline{\sum\left(\sum\left(\sum D_n+F_m\right)+\sigma_l^d\right)}}\,\mathsf{app}}{\mathcal{E}+\sum\left(\sum\left(\sum D_n+F_m\right)+\sigma_l^d\right)}\,\mathsf{uni}}{\mathcal{E}+\psi}\,\mathsf{del}$$

- obs/del-app-**reduction (2)**
When $\overrightarrow{\varphi_n}, \overrightarrow{\mathcal{F}'_m}, \overrightarrow{\sigma_l^d} \vdash \mathcal{E}$ and $\sum\left(\sum\left(\sum D_n+F_m\right)+\sigma_l^d\right)+\psi$ is defined in the following part in a proof:

$$\begin{array}{c}\begin{array}{c}\begin{array}{cccccc}\vdots\,\pi_1^2 & \vdots\,\pi_m^2 & \vdots\,\pi_1^1 & \vdots\,\pi_n^1 & \vdots\,\pi_1^3 & \vdots\,\pi_l^3\\ \dfrac{\mathcal{F}_1}{\mathcal{F}'_1}\,\mathsf{del}\cdots & \dfrac{\mathcal{F}_m}{\mathcal{F}'_m}\,\mathsf{del}\cdots & \dfrac{\mathcal{D}_1}{\varphi_1}\,\mathsf{obs}\cdots & \dfrac{\mathcal{D}_n}{\varphi_n}\,\mathsf{obs}\cdots & \sigma_1^d & \cdots\sigma_l^d\end{array}\\ \vdots\,\pi'\\ \mathcal{E}\end{array}\qquad\qquad\begin{array}{c}\vdots\,\pi''\\ \psi\end{array}\\ \hline \mathcal{E}+\psi\end{array}\;\mathsf{app}$$

where $m \neq 0$ or $n \neq 0$, it is reduced to:

$$\begin{array}{c}\begin{array}{c}\begin{array}{c}\begin{array}{c}\begin{array}{c}\begin{array}{cc}\vdots\,\pi_1^1 & \vdots\,\pi_n^1\\ \mathcal{D}_1 & \cdots\mathcal{D}_n\end{array}\\ \hline\hline \sum\mathcal{D}_n\end{array}\;\mathsf{uni}\quad\begin{array}{cc}\vdots\,\pi_1^2 & \vdots\,\pi_m^2\\ \mathcal{F}_1 & \cdots\mathcal{F}_m\end{array}\\ \hline\hline \sum\left(\sum\mathcal{D}_n+\mathcal{F}_m\right)\end{array}\;\mathsf{uni}\quad\begin{array}{cc}\vdots\,\pi_1^3 & \vdots\,\pi_l^3\\ \sigma_1^d & \cdots\sigma_l^d\end{array}\\ \hline\hline \sum\left(\sum\left(\sum D_n+F_m\right)+\sigma_l^d\right)\end{array}\;\mathsf{app}\quad\begin{array}{c}\vdots\,\pi''\\ \psi\end{array}\\ \hline \sum\left(\sum\left(\sum D_n+F_m\right)+\sigma_l^d\right)+\psi\end{array}\;\mathsf{app}\\ \hline \mathcal{E}+\psi\end{array}\;\mathsf{del}$$

- del/obs-uni-**reduction (1)**

When $\overrightarrow{\varphi_n}, \overrightarrow{\mathcal{F}'_m}, \overrightarrow{\sigma_l^d} \vdash \mathcal{D}$ and $\sum\left(\sum\left(\sum D_n + F_m\right) + \sigma_l^d\right) + \mathcal{E}$ is defined in the following part in a proof:

$$\begin{array}{c}\begin{array}{c}\begin{array}{cccccc}\vdots\,\pi_1^1 & \vdots\,\pi_n^1 & \vdots\,\pi_1^2 & \vdots\,\pi_m^2 & \vdots\,\pi_1^3 & \vdots\,\pi_l^3\\ \dfrac{\mathcal{D}_1}{\varphi_1}\,\mathsf{obs}\cdots & \dfrac{\mathcal{D}_n}{\varphi_n}\,\mathsf{obs}\cdots & \dfrac{\mathcal{F}_1}{\mathcal{F}'_1}\,\mathsf{del}\cdots & \dfrac{\mathcal{F}_m}{\mathcal{F}'_m}\,\mathsf{del}\cdots & \sigma_1^d & \cdots\sigma_l^d\end{array}\\ \vdots\,\pi'\\ \mathcal{D}\end{array}\qquad\qquad\begin{array}{c}\vdots\,\pi''\\ \mathcal{E}\end{array}\\ \hline \mathcal{D}+\mathcal{E}\end{array}\;\mathsf{uni}$$

where $m \neq 0$ or $n \neq 0$, it is reduced to:

$$\begin{array}{c}\begin{array}{c}\begin{array}{c}\begin{array}{c}\begin{array}{c}\begin{array}{cc}\vdots\,\pi_1^1 & \vdots\,\pi_n^1\\ \mathcal{D}_1 & \cdots\mathcal{D}_n\end{array}\\ \hline\hline \sum\mathcal{D}_n\end{array}\;\mathsf{uni}\quad\begin{array}{cc}\vdots\,\pi_1^2 & \vdots\,\pi_m^2\\ \mathcal{F}_1 & \cdots\mathcal{F}_m\end{array}\\ \hline\hline \sum\left(\sum\mathcal{D}_n+\mathcal{F}_m\right)\end{array}\;\mathsf{uni}\quad\begin{array}{cc}\vdots\,\pi_1^3 & \vdots\,\pi_l^3\\ \sigma_1^d & \cdots\sigma_l^d\end{array}\\ \hline\hline \sum\left(\sum\left(\sum D_n+F_m\right)+\sigma_l^d\right)\end{array}\;\mathsf{app}\quad\begin{array}{c}\vdots\,\pi''\\ \mathcal{E}\end{array}\\ \hline \sum\left(\sum\left(\sum D_n+F_m\right)+\sigma_l^d\right)+\mathcal{E}\end{array}\;\mathsf{uni}\\ \hline \mathcal{D}+\mathcal{E}\end{array}\;\mathsf{del}$$

- del/obs-uni-**reduction (2)** is defined similarly for $\pi''$.

A **redex** is a tuple of applications of rules and diagrammatic formulas (obs, $\ldots$, obs, del, $\ldots$, del, $\sigma_1^d, \ldots, \sigma_l^d$; app) or (del, $\ldots$, del, obs, $\ldots$, obs, $\sigma_1^d, \ldots, \sigma_l^d$; uni) to which, a reduction rule can be applied. A heterogeneous proof is said to be in **normal form** when it does not contain any redex.

*Example 2 (*obs-app*-reduction).* By reducing the obs-app pair of the following proof on the left in Fig. 1, we obtain the normal proof on the right.

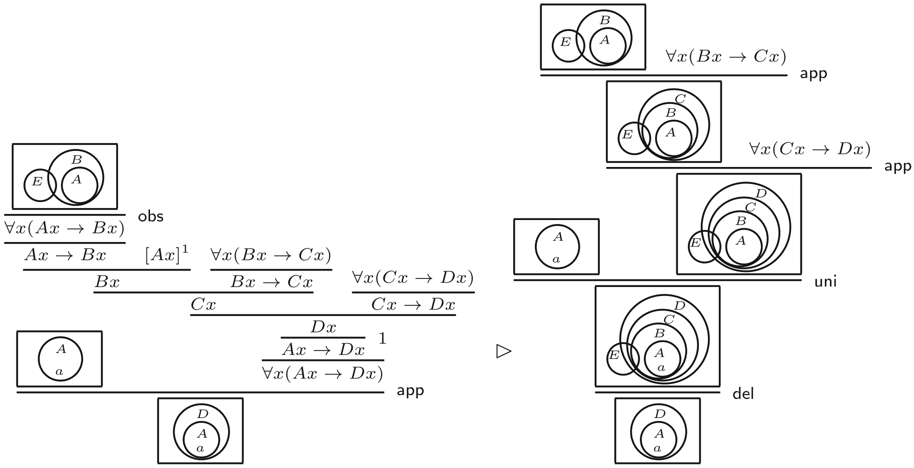


**Fig. 1.** obs-app-reduction

Note that in a given proof, our redex is not uniquely determined, and there are several choices of a redex with respect to the same app/uni. Cf. Fig. 2.

### 4.3 Normalization

Based on our reduction rules, we establish our normalization theorem. One of the difficulties is that after an obs/del-app-reduction, new applications of app and uni are provided and they may induce new redexes. To overcome this difficulty, we choose the topmost-leftmost redex in a given proof, and apply our reduction twice in a row.

**Theorem 1 (Normalization).** *Let $\alpha_i$ be a diagram or a formula. Any proof of $\alpha$ from $\alpha_1, \ldots, \alpha_n$ is reduced to a normal proof of $\alpha$ from $\alpha_1, \ldots, \alpha_n$.*

*Proof.* We distinguish linguistic FOL parts and diagrammatic parts in a given proof, and we first reduce the linguistic parts, whose normalization theorem is already established. Let $\pi$ be a heterogeneous proof whose linguistic parts are already reduced to normal form. For every application of app or uni in $\pi$, we define its degree $deg(\mathsf{app})$ or $deg(\mathsf{uni})$ as the number of applications of obs and del that form redexes with respect to the application of app or uni in question. Thus, in any normal form, $deg(\mathsf{app}) = deg(\mathsf{uni}) = 0$ for every application of app and uni in the proof. We choose the topmost-leftmost redex in $\pi$, which is the leftmost application of app or uni whose degree is minimal in $\pi$. We divide the cases according to the rule: app or uni.

When the rule is app, let the topmost-leftmost redex be the following form:

$$\dfrac{\begin{array}{c}\dfrac{\overset{\vdots\;\pi_1^1}{\mathcal{E}_1}}{\mathcal{E}_1'}\,\mathsf{del}\cdots\dfrac{\overset{\vdots\;\pi_k^1}{\mathcal{E}_k}}{\mathcal{E}_k'}\,\mathsf{del}\\ \vdots\;\pi'\\ \mathcal{E}\end{array}\qquad\begin{array}{c}\dfrac{\overset{\vdots\;\pi_1^2}{\mathcal{D}_1}}{\varphi_1}\,\mathsf{obs}\cdots\dfrac{\overset{\vdots\;\pi_n^2}{\mathcal{D}_n}}{\varphi_n}\,\mathsf{obs}\cdots\dfrac{\overset{\vdots\;\pi_1^3}{\mathcal{F}_1}}{\mathcal{F}_1'}\,\mathsf{del}\cdots\dfrac{\overset{\vdots\;\pi_m^3}{\mathcal{F}_m}}{\mathcal{F}_m'}\,\mathsf{del}\cdots\overset{\vdots\;\pi_1^4}{\sigma_1^d}\cdots\overset{\vdots\;\pi_l^4}{\sigma_l^d}\\ \vdots\;\pi''\\ \psi\end{array}}{\mathcal{E}+\psi}\,\mathsf{app}$$

where every del and obs is the topmost application that forms a redex with respect to the given app, and hence there are no obs nor del in $\overrightarrow{\pi_k^1}, \overrightarrow{\pi_n^2}, \overrightarrow{\pi_m^3}$ that forms a redex with respect to the given app.

Note that there is application of neither obs nor del in $\overrightarrow{\pi_l^4}$. Since, if such an application exists, it has to be one of $\pi^2$ or $\pi^3$. Furthermore, above $\mathcal{E}$, i.e., in $\pi'$ and $\overrightarrow{\pi_k^1}$, there is only application of del without obs, or there is application of neither del nor obs (i.e., $k = 0$). This is because, if there is an obs, then there has to be an application of app for inferring the diagram $\mathcal{E}$ and $\mathcal{E}_i$. This contradicts the assumption that the given app forms the topmost-leftmost redex.

By applying del-app-reduction to $\pi'$, we obtain the following proof:

$$\dfrac{\dfrac{\begin{array}{c}\overset{\vdots\;\pi_1^1}{\mathcal{E}_1}\cdots\overset{\vdots\;\pi_k^1}{\mathcal{E}_k}\\ \hline\hline \sum\mathcal{E}_k\end{array}\,\mathsf{uni}\qquad\begin{array}{c}\dfrac{\overset{\vdots\;\pi_1^2}{\mathcal{D}_1}}{\varphi_1}\,\mathsf{obs}\cdots\dfrac{\overset{\vdots\;\pi_n^2}{\mathcal{D}_n}}{\varphi_n}\,\mathsf{obs}\cdots\dfrac{\overset{\vdots\;\pi_1^3}{\mathcal{F}_1}}{\mathcal{F}_1'}\,\mathsf{del}\cdots\dfrac{\overset{\vdots\;\pi_m^3}{\mathcal{F}_m}}{\mathcal{F}_m'}\,\mathsf{del}\cdots\overset{\vdots\;\pi_1^4}{\sigma_1^d}\cdots\overset{\vdots\;\pi_l^4}{\sigma_l^d}\\ \vdots\;\pi''\\ \psi\end{array}}{\sum\mathcal{E}_k+\psi}\,\mathsf{app}}{\mathcal{E}+\psi}\,\mathsf{del}$$

By further applying obs/del-app-reduction to $\pi''$, we obtain the following proof:

$$\dfrac{\dfrac{\begin{array}{c}\overset{\vdots\;\pi_1^1}{\mathcal{E}_1}\cdots\overset{\vdots\;\pi_k^1}{\mathcal{E}_k}\\ \hline\hline \sum\mathcal{E}_k\end{array}\,\mathsf{uni}\qquad\begin{array}{c}\begin{array}{c}\begin{array}{c}\overset{\vdots\;\pi_1^2}{\mathcal{D}_1}\cdots\overset{\vdots\;\pi_n^2}{\mathcal{D}_n}\\ \hline\hline \sum\mathcal{D}_n\end{array}\,\mathsf{uni}\quad\overset{\vdots\;\pi_1^3}{\mathcal{F}_1}\cdots\overset{\vdots\;\pi_m^3}{\mathcal{F}_m}\\ \hline\hline \sum\left(\sum\mathcal{D}_n+\mathcal{F}_m\right)\end{array}\,\mathsf{uni}\quad\overset{\vdots\;\pi_1^4}{\sigma_1^d}\cdots\overset{\vdots\;\pi_l^4}{\sigma_l^d}\\ \hline\hline \sum\left(\sum\left(\sum\mathcal{D}_n+\mathcal{F}_m\right)+\sigma_l^d\right)\end{array}\,\mathsf{app}}{\sum\mathcal{E}_k+\sum\left(\sum\left(\sum\mathcal{D}_n+\mathcal{F}_m\right)+\sigma_l^d\right)}\,\mathsf{uni}}{\mathcal{E}+\psi}\,\mathsf{del}$$

Although new applications of uni and app are produced, there is neither obs nor del in $\overrightarrow{\pi_k^1}, \overrightarrow{\pi_n^2}, \overrightarrow{\pi_m^3}, \overrightarrow{\pi_l^4}$ that forms a new redex with respect to these uni and app. This is justified because, if there is an obs or del that forms a new redex, then it must have already formed a redex in the original proof, which contradicts the assumption that the deleted obs and del are topmost. Thus, the degrees of these new uni and app are 0.

The case where the topmost-leftmost redex consists of an application of uni is similar. Let the degree $deg(\pi)$ of the given proof $\pi$ be the sum of all degrees of applications of app and uni in $\pi$. Let $\pi^{\star}$ be the proof obtained by the above topmost-leftmost reduction. Then, we have $deg(\pi^{\star}) < deg(\pi)$. Therefore, by repeated applications of the topmost-leftmost reduction, we obtain a proof whose degree is 0, i.e., we obtain a normal proof. ■

*Example 3 (Normalization).* By repeatedly reducing the topmost-leftmost redex, we obtain a normal proof as shown in Fig. 2.

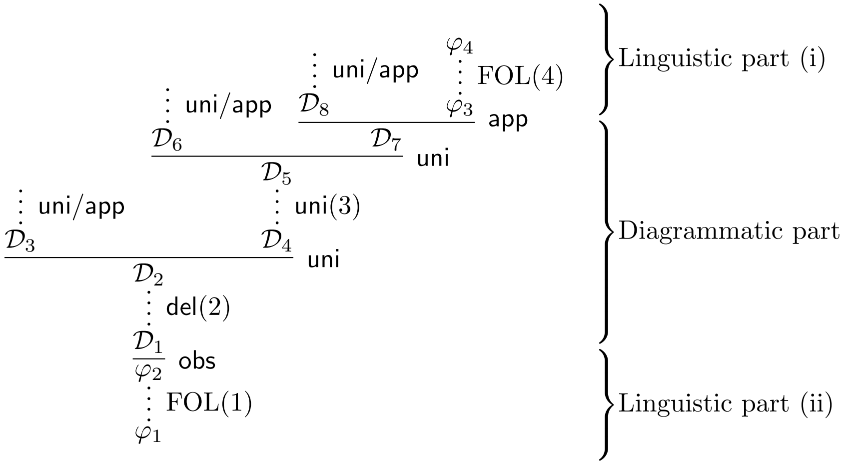
Note that in the normal form, the application of uni seems to be redundant, since one of its premises and its conclusion are the same diagrams. There may be this kind of redundancy in our normal form, as we see in the usual normal form in natural deduction for FOL. However, this kind of redundancy is different from our essential detour, which uses redundant subconcepts. Thus, we leave this kind of inessential redundant parts untouched in our normal form.

By applying our reduction, every redundant linguistic part that lies between diagrammatic parts is removed. From the perspective of diagrams reducing certain complexity of linguistic inference, it is ineffective to infer diagrammatically by way of some linguistic parts, and hence our reduction is also verified from this perspective.

## 4.4 Characterization of Normal Heterogeneous Proofs

Let us investigate how diagrammatic inference and linguistic inference appear, and are related in our heterogeneous proofs. The following proposition holds straightforwardly in a system that does not constrain any inference rules such as Venn and Euler systems without any points.

**Proposition 1 (Normal form).** *In a heterogeneous system, where* app *and* uni *are applicable to any diagram and formula without any constraint, every normal proof has the following form:*



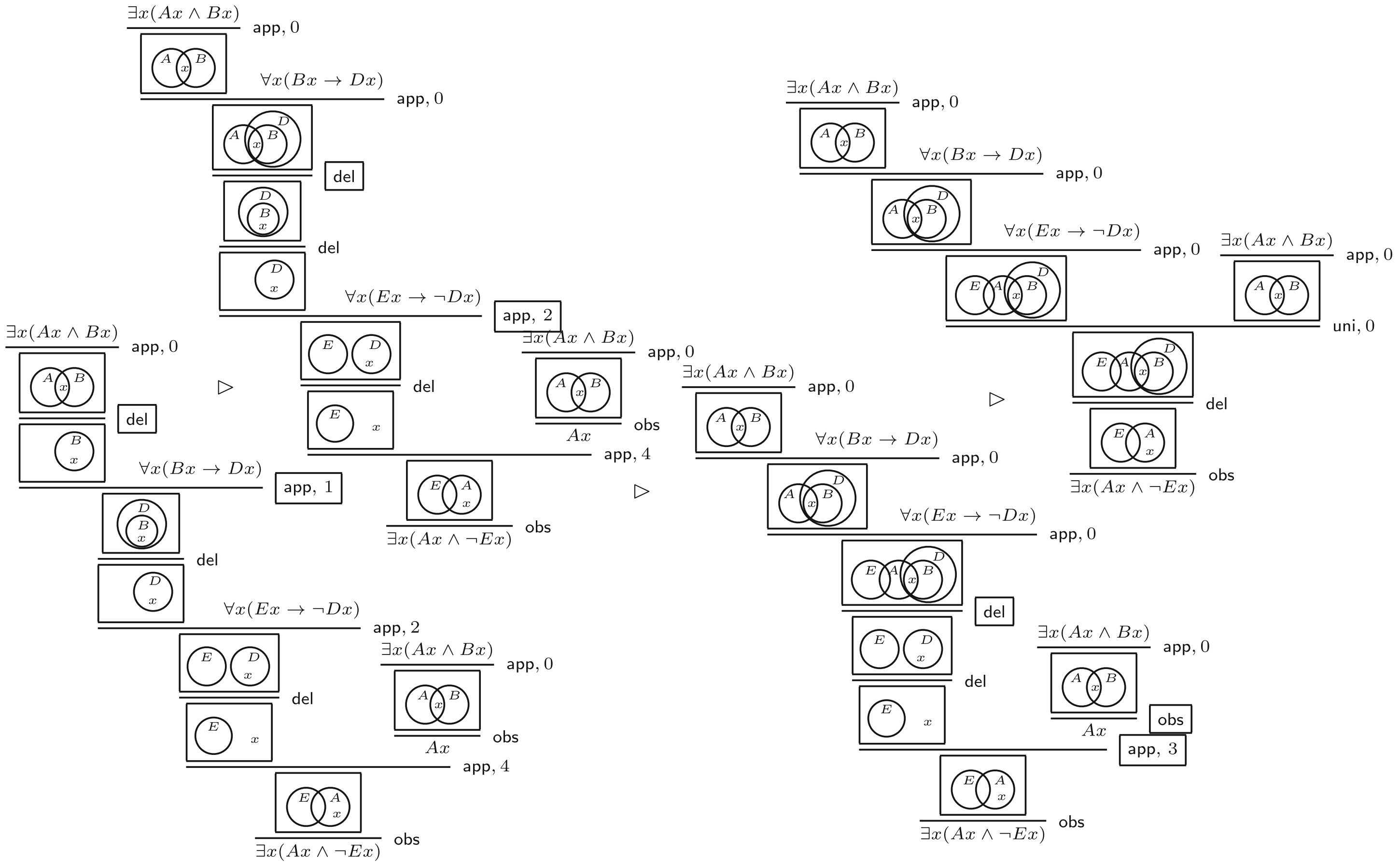


**Fig. 2.** Normalization. (The number beside an application of app or uni is its degree.)

*Proof.* Starting from a given conclusion, we examine possible rules in a normal proof from the bottom-up. The following number corresponds to that in the above figure. Assume that the given conclusion is a formula $\varphi_1$. When the conclusion is a diagram, we start from the following case (2).

(1) $\varphi_1$ should be obtained by FOL-rules or obs, since the conclusion of other rules del,uni, and app is a diagram. Thus, we assume without loss of generality, that $\varphi_1$ is obtained after application of FOL-rules followed by an application of obs from the bottom-up.
(2) Above the obs of (1), possible rules are del, uni, and app. (FOL-rules and obs are not possible, since their conclusion has to be a formula.) Since there is no del above uni or app in a normal proof, we assume without loss of generality, that del is applied certain times.
(3) Above the del-rule of (2), possible rules are uni or app. Although any of them is possible, we assume without loss of generality, that $\mathcal{D}_2$ is obtained by some applications of uni followed by an application of app.
(4) Above the app of (3), the only possible rule is one of the FOL-rules, since there is no obs (nor del) above app in a normal proof. Therefore, only successive applications of FOL-rules are possible. ■

By Proposition 1 above, a normal heterogeneous proof is divided into the following three parts from the top-down, which also indicates a strategy to construct heterogeneous proofs.

**Linguistic part (i)** By FOL-rules, given premises represented by formulas are decomposed, and apply-formulas are constructed.

**Diagrammatic part** uni is applied to given premises represented by diagrams, and app is applied to apply-formulas obtained at the linguistic part (i), and a maximal diagram is constructed. Then, by del and obs, diagrammatic formulas are extracted from the maximal diagram.

**Linguistic part (ii)** By FOL-rules, the conclusion is constructed.

## 5 Discussion and Future Work

By slightly extending the notion of free ride of Shimojima [13], let us call diagrammatic formulas of the conclusion of app and uni, *free rides*, if they do not appear in the given premise diagrams or sentences, but (automatically) appear in its conclusion [15]. From the perspective of symbolic specification, a diagram is a deductively closed set of diagrammatic formulas. The deductive closedness of diagrams induces the free rides. The larger a diagram is, the more free rides appear in general. Thus, since a maximal diagram is constructed in our normal heterogeneous proof, we may say that a normal heterogeneous proof takes full advantage of the free rides. However, from a cognitive standpoint, a maximal diagram is not necessarily comprehensible or manageable. This is because, the more complex a diagram is, more cognitive cost is required to construct and read the diagram in general. In [8], to make proofs readable by avoiding clutter in diagrams, tactics are introduced to an interactive theorem prover for spider diagrams Speedith [18].

It is often pointed out that there is a trade-off between the expressive power and the cognitive clarity/complexity of diagrams. In general, on top of inherent geometrical constraints of diagrams, if we increase their expressive power by introducing various conventional devices (for example, linking between points as well as between diagrams), then it is appropriate that the cognitive clarity of the diagrams is decreased. Conversely, if we restrict introducing conventional devices, then such diagrams maintain their cognitive clarity in exchange for limited expressive power. Our characterization of heterogeneous proofs of Proposition 1 shows another trade-off between constraints on inference rules, and the complexity of the structure of proofs or of the strategy to construct proofs. As discussed in Sect. 3.2, there may be various constraints on inference rules such as constraints for indeterminacy and for contradiction of `app` and `uni`. Such a constraint is mainly imposed to avoid cognitive complexity or to maintain actual feasibility of the rule. (Although a constraint on inference rules pertains to expressive limitation of diagrams, they are not the same.) The characterization of the structure for heterogeneous proofs of Proposition 1 is valid for systems without constraints on inference rules `app` and `uni`. Thus, in a system with some constraints, we cannot apply our strategy to construct proofs in a straightforward manner. We need a more complex strategy or a heuristic method to construct proofs in such a system. In general, within a system with various constraints on inference rules, although cognitive clarity and actual feasibility of the rules are maintained due to those constraints, the structure of proofs in such a system becomes complex and an automatic strategy to construct proofs cannot be applied. Conversely, in a system with few constraints on rules, the structure of proofs in such a system is simpler and we are able to apply an automatic strategy to construct proofs. However, the cognitive clarity and actual feasibility of each inference rule are decreased in such a system.

In this article, we restrict our type of a diagram to the conjunction of diagrammatic formulas thereof, and hence, we exclude from our consideration devices to express disjunctive information. For future work, we aim to extend our framework in order to include such diagrams representing disjunctive information.

## References

1. Barker-Plummer, D., Barwise, J., Etchemendy, J.: Logical Reasoning with Diagrams & Sentences: Using Hyperproof. CSLI Publications, Stanford (2017)
2. Barker-Plummer, D., Swoboda, N.: Reasoning with coincidence grids–a sequent-based logic and an analysis of complexity. J. Vis. Lang. Comput. **22**(1), 56–65 (2011)
3. Barwise, J., Etchemendy, J.: Hyperproof: For Macintosh. The Center for the Study of Language and Information Publications, Stanford (1995)
4. van Dalen, D.: Logic and Structure, 4th edn. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-85108-0

5. Gentzen, G.: Untersuchungen über das logische Schließen, Mathematische Zeitschrift, **39**, 176–210, 405–431 (1934). English Translation: Investigations into logical deduction. In: Szabo, M.E. (ed.) The Collected Papers of Gerhard Gentzen (1969)
6. Hammer, E.: Reasoning with sentences and diagrams. Notre Dame J. Form. Log. **35**(1), 73–87 (1994)
7. Howse, J., Stapleton, G., Taylor, J.: Spider diagrams. LMS J. Comput. Math. **8**, 145–194 (2005). London Mathematical Society
8. Linker, S., Burton, J., Jamnik, M.: Tactical diagrammatic reasoning. In: Proceedings of 12th Workshop on User Interfaces for Theorem Provers, EPTCS, vol. 239, pp. 29–42 (2017)
9. Mineshima, K., Okada, M., Takemura, R.: A diagrammatic inference system with Euler circles. J. Log. Lang. Inform. **21**(3), 365–391 (2012)
10. Prawitz, D.: Natural Deduction. Almqvist & Wiksell, Stockholm (1965). (Dover 2006)
11. Prawitz, D.: Ideas and results in proof theory. In: Proceedings of 2nd Scandinavian Logic Symposium, pp. 237–309 (1971)
12. Shin, S.-J.: The Logical Status of Diagrams. Cambridge University Press, Cambridge (1994)
13. Shimojima, A.: Semantic Properties of Diagrams and Their Cognitive Potentials. CSLI Publications, Stanford (2015)
14. Swoboda, N., Allwein, G.: Heterogeneous reasoning with Euler/Venn diagrams containing named constants and FOL. Electron. Notes Theoret. Comput. Sci. **134**, 153–187 (2005)
15. Takemura, R.: Proof theory for reasoning with Euler diagrams: a logic translation and normalization. Stud. Logica. **101**(1), 157–191 (2013)
16. Takemura, R., Shimojima, A., Katagiri, Y.: A logical investigation on global reading of diagrams. In: Cox, P., Plimmer, B., Rodgers, P. (eds.) Diagrams 2012. LNCS (LNAI), vol. 7352, pp. 330–333. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31223-6_42
17. Urbas, M., Jamnik, M.: Heterogeneous proofs: spider diagrams meet higher-order provers. In: van Eekelen, M., Geuvers, H., Schmaltz, J., Wiedijk, F. (eds.) ITP 2011. LNCS, vol. 6898, pp. 376–382. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22863-6_29
18. Urbas, M., Jamnik, M., Stapleton, G.: Speedith: a reasoner for spider diagrams. J. Log. Lang. Inform. **24**(4), 487–540 (2015)

# Iconic Logic and Ideal Diagrams: The Wittgensteinian Approach

Timm Lampert(✉)

Department of Philosophy, Humboldt University Berlin, Berlin, Germany
`lampertt@staff.hu-berlin.de`

**Abstract.** This paper provides a programmatic overview of a conception of iconic logic from a Wittgensteinian point of view (WIL for short). The crucial differences between WIL and a standard version of symbolic logic (SSL) are identified and discussed. WIL differs from other versions of logic in that in WIL, logical forms are identified by means of so-called ideal diagrams. A logical proof consists of an equivalence transformation of formulas into ideal diagrams, from which logical forms can be read off directly. Logical forms specify properties that identify sets of models (conditions of truth) and sets of counter-models (conditions of falsehood). In this way, WIL allows the sets of models and counter-models to be described by finite means. Against this background, the question of the decidability of first-order-logic (FOL) is revisited. In the last section, WIL is contrasted with Peirce's iconic logic (PIL).

## 1 Introduction

This paper outlines an alternative to standard symbolic logic (SSL), namely, Wittgenstein's iconic logic (WIL), as a basis for first-order logic (FOL), while avoiding the algorithmic details.[1]

I call the outlined approach "Wittgensteinian" for two reasons: (i) it is inspired by Wittgenstein's early philosophy of logic, and (ii) I wish to distinguish it from Peirce's conception of an iconic logic (PIL). However, I will not present any justification demonstrating that the outlined conception of logic is indeed that of Wittgenstein's early works, nor will I compare the details of Wittgenstein's and Peirce's approaches. Instead, I will focus on the programmatic ideas and fundamental concepts of this Wittgensteinian approach to iconic logic (WIL). In doing so, I intend (i) to make manifest that FOL can be pursued within different paradigms and (ii) to encourage others to work within a Wittgensteinian paradigm of iconic logic.

---

T. Lampert—I am grateful to Wulf Rehder for many helpful comments on an earlier draft of this paper.

[1] Algorithms that realize some of Wittgenstein's ideas concerning logical proofs are available at the following link: http://www2.cms.hu-berlin.de/newlogic/webMathematica/Logic/home.jsp.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 624–639, 2018.
https://doi.org/10.1007/978-3-319-91376-6_56

I begin by given the rationale behind WIL (Sect. 2). In the main body of the paper, I explain the conception of proof in WIL and the crucial notion of ideal diagrams as representations of logical forms (Sects. 3 to 5). I then allude to several significant differences that arise when applying WIL and SSL by addressing the questions of adequate formalization and decidability (Sects. 6 and 7). Finally, I distinguish WIL from PIL (Sect. 8).

Since the concepts of *logical forms* and *ideal diagrams* are crucial, I will define them here at the outset. Concrete examples and explanations of the concepts used in these definitions will be given below in Sects. 2 to 5.

**Logical Form:** The logical form of a first-order formula $\phi$ is the form of the conditions for truth and falsehood that hold for all formulas that are logically equivalent to $\phi$.

According to WIL, ideal diagrams represent logical forms unambiguously. Ideal diagrams are unique representations of equivalence classes of logical formulas. I define them by using (i) a pole-group notation that Wittgenstein introduced in his early writings and (ii) minimal disjunctive normal forms of first-order logic (minimal FOLDNFs). The complete details will be presented in Sects. 4 and 5.

**Ideal Diagram:** An ideal diagram is the translation of the set of minimal FOLDNFs that is generated from an initial formula $\phi$ into Wittgenstein's pole-group notation.

*Paraphrases of ideal diagrams*, in turn, are the results of a mechanical reading algorithm for ideal diagrams. They make use of a standardized informal language that makes explicit how ideal diagrams should be read as representations of the conditions for the truth and falsehood of instances of initial formulas.

## 2 The Case for the WIL Approach

Russell writes the following in [Russell (1992)], p. xvi:

> The fundamental characteristic of logic, obviously, is that which is indicated when we say that logical propositions are true in virtue of their form. [...] I confess, however, that I am unable to give any clear account of what is meant by saying that a proposition is "true in virtue of its form".

In SSL, "logical propositions" are defined as formulas that are true in all interpretations. In this sense, SSL places priority on semantics. Accordingly, it does not make sense to characterize logical propositions as "true in virtue of their form". The set of logical propositions is defined not by any specific logical form shared by all logical propositions but rather by the characteristic of being true in any interpretation. In the case of FOL, this means that logical propositions cannot be identified algorithmically by evaluating single interpretations because the number of possible interpretations is infinite.

By contrast, WIL can be characterized as a logic that is intended to fulfill Russell's desideratum. In general, the primary aim of WIL is to assign logical forms to equivalence classes of logical formulas. It is important to recognize that such a conception is reasonable only if one does not rely on either paraphrases or interpretations based on the structure of the *logical formulas* whose logical properties are in question. Such methods of reading or evaluating formulas do not refer to anything that is common to all formulas in the same set of logically equivalent formulas and that may thus serve to identify conditions for the truth or falsehood of propositions sharing the same logical form.

According to [Etchemendy (1999)], there are two ways of understanding the semantics of a formal language. In the *representational view*, different models and counter-models represent different *logically possible* configurations of the world. According to this view, "interpretations" are understood as conditions for the truth value of a sentence. Instances of propositional function variables are fixed, and their meanings do not change with varying interpretations; only their truth values do. By contrast, in the *interpretational view*, different models and counter-models correspond to the assignment of different actual extensions to expressions. This conception does not consider "logical possibilities" or "meaning" in terms of conditions for truth and falsehood. The interpretational view is the standard view of mathematical logic, for example, in Tarski's semantics. The representational view, by contrast, is commonly adopted in philosophical approaches to the semantics of FOL. WIL essentially adopts this view; hence, referring to models and counter-models is equivalent to referring to conditions for truth and falsehood in terms of various logically possible states of the world. According to WIL, the general task of logic is to distinguish conditions for truth and falsehood within a space of *logical possibilities* by identifying the logical form of admissible instances of logical formulas.

In WIL, the logical form of a formula must first be revealed, and it is not until such a logical form has been identified that one can answer the question of what such a form contributes to the representation of conditions for truth and falsehood. As in the case of ordinary propositions, the outer form of a logical formula disguises its logical form. This is so for the following reasons:

1. Any set of logically equivalent formulas is infinite, and although all of the equivalent formulas in such a set share the same logical form, they may have different outer forms. For example, although formulas such as $P$, $P \vee P$, $P \vee Q \wedge \neg Q$ and $P \vee \neg(R \vee \neg R)$ differ from each other, instances of these different formulas share the same conditions for truth and falsehood.
2. Consequently, one cannot paraphrase an arbitrary logical formula such that
   (a) the paraphrase clarifies what each sign contributes to the representation of the conditions for truth and falsehood (i.e., how each part of the formula specifies certain properties of models or counter-models),
   (b) the signs are *unambiguously* paraphrased to achieve such a clarification (i.e., identical signs are paraphrased identically and different signs are paraphrased differently), and
   (c) all of the (finite number of) *non-redundant* paraphrases of the conditions for truth and falsehood are provided (i.e., all paraphrases that do not contain any part that can be eliminated without resulting in a paraphrase of a different set of models or counter-models).

By contrast, in WIL, one and only one ideal diagram is assigned to all equivalent formulas, and a proper reading algorithm for such ideal diagrams satisfies conditions 2(a) to 2(c). In doing so, such an algorithm "reads off" the logical form from an ideal diagram.

WIL qualifies as an "iconic logic" because ideal diagrams identify logical forms by their syntactic properties. The features of ideal diagrams serve as identity criteria for sets of (counter-)models that share certain properties. Syntax is prior to semantics in WIL in the sense that for a given formula, the properties of models and counter-models are identified prior to and independently of the evaluation of that formula with respect to single interpretations.

According to WIL, not only the outer form of ordinary language but also the outer form of logical formulas can lead to (logical, linguistic or philosophical) misunderstandings. WIL avoids such misunderstandings by revealing logical forms through equivalence transformation. Such a procedure elucidates our implicit understanding of the construction of logical formulas and what it contributes to specifying conditions for the truth and falsehood of propositions.

## 3 Logical Proofs

In SSL, logical proofs derive theorems from axioms (or auxiliary assumptions) within a correct and complete calculus. In WIL, however, a proof procedure transforms initial logical formulas into ideal diagrams that enable the identification of the corresponding logical form. Hence, logical proofs in WIL are not merely proofs of logical theorems. A proof in WIL answers the more general question of how an initial formula contributes to identifying conditions for truth and falsehood in general. The proof of a logical theorem (or, likewise, a logical contradiction) is merely a special case of this general procedure.

Because ideal diagrams identify conditions for truth and falsehood and, consequently, also allow one to decide whether the initial formulas are "true in all interpretations", a proof procedure in WIL amounts to a decision procedure. I will discuss the general question of decidability in Sect. 7. For now, it may suffice to say that the crucial challenge in WIL is to specify algorithms for transforming logical formulas into ideal diagrams. In the remainder of this paper, I will present a programmatic overview of WIL, without discussing the technical details of the algorithms for solving this problem. In the following two sections, however, I will address the question of how to specify the ideal diagrams that result from the aforementioned transformation from initial formulas.

## 4 Ideal Diagrams I - Propositional Logic

From 1912 to 1914, Wittgenstein developed his so-called *ab*-notation as a means of uniquely representing conditions for the truth and falsehood of propositions of a certain logical form.[2] He illustrated this notation with various diagrams

---

[2] Cf. his letters to Russell during this period, reproduced in [Wittgenstein (1997)], as well as Wittgenstein's *Notes on Logic* and his *Notes dictated to G.E. Moore*, both printed in [Wittgenstein (1979)].

of several logical formulas. He used similar diagrams in [Wittgenstein (1994)], remark 6.1203, to demonstrate how to identify tautologies by applying syntactic criteria to the resulting expressions. Instead of the *ab*-notation, he used $T$ and $F$ as "poles" representing the possibilities of truth and falsehood. Wittgenstein also suggested transforming his diagrams into a simpler pole-group notation that corresponds to certain disjunctive normal forms (DNFs) (cf. [Wittgenstein (1979)], p. 102, and [Wittgenstein (1997)], letter 30). His notation was intended to apply not only to propositional logic but also to FOL (cf. [Wittgenstein (1979)], p. 95f). In a letter to Russell, he even conjectured that applying his notation to FOL would enable the identification of tautologies throughout the entire realm of FOL (cf. [Wittgenstein (1997)], letter 30). However, he never spelled out in detail how to apply his notation to arbitrary FOL formulas, nor did he discuss in detail how to achieve unique representations of logical forms in propositional logic (or even FOL). The following is an attempt to revisit Wittgenstein's claim and specify in more detail what is needed in order to represent logical forms by means of ideal diagrams. In this short paper, I cannot elaborate all of the rules for generating such diagrams from logical formulas. Instead, I will focus only on their general properties.

I will initially restrict the discussion to propositional logic. In this case, the application of the well-known Quine-McCluskey algorithm to obtain a set of minimal DNFs is a crucial step in the generation of ideal diagrams. Minimal DNFs distinguish sufficient conditions for truth (the disjuncts) and non-redundant parts of those conditions (the conjuncts); cf. condition 2(a) on p. 3. This allows conditions for the truth of admissible instances of an initial formula to be read off. The same applies to conditions for falsehood, if one also generates the set of minimal DNFs of the negation of the initial formula. By the nature of minimal DNFs, no part of the paraphrase of any single minimal DNF is redundant; cf. condition 2(c) on p. 3.

However, the minimal DNFs of a formula of propositional logic are not unique. Therefore, their paraphrase does not satisfy condition 1 on p. 3. For example, formula (1) has the two minimal DNFs expressed in (2) and (3):

$$P \wedge \neg Q \vee \neg P \wedge Q \vee P \wedge R \vee Q \wedge R \tag{1}$$

$$P \wedge \neg Q \vee \neg P \wedge Q \vee P \wedge R \tag{2}$$

$$P \wedge \neg Q \vee \neg P \wedge Q \vee Q \wedge R \tag{3}$$

However, if one regards a representation of the entire finite set of minimal DNFs as the ideal diagram, then the requirement of uniqueness is satisfied. One might object that if both formulas (2) and (3) together are taken to be part of the ideal diagram, then the non-redundancy requirement for the paraphrases of ideal diagrams (cf. condition 2(c) on p. 3) is not satisfied. However, I propose to interpret the non-uniqueness of the minimal DNFs in terms of an "ambiguity of the logical form". This ambiguity is represented by a corresponding ambiguity within the ideal diagram. Therefore, the ideal diagram must represent all alternative minimal DNFs, and thus, no alternative is superfluous. Each alternative might be called a representation of a "variant" of the logical form. Such an alternative must not, in itself, contain any redundancy in the description of the conditions for truth or falsehood. However, all alternatives in the entire set of such alternatives must be considered in order to characterize the ambiguity of the logical form. The extent of that ambiguity can be quantified by the number of minimal DNFs. This is why condition 2(c) on p. 3 refers to "*all* non-redundant paraphrases". The non-redundancy requirement applies only to each paraphrase individually and not to the set of all such paraphrases.

Instead of representing all minimal DNFs as a set, one may distinguish components that are common to all minimal DNFs from those that are different by using a two-dimensional notation as in expression (4):

$$P \wedge \neg Q \vee \neg P \wedge Q \vee \genfrac{}{}{0pt}{2}{P\wedge R}{Q\wedge R} \tag{4}$$

In contrast to (4), only one minimal DNF is generated from ¬(1) to represent the conditions for the falsehood of propositions that are admissible instances of (1):

$$\neg P \wedge \neg Q \vee P \wedge Q \wedge \neg R \tag{5}$$

Wittgenstein's *ab*- or `TF`-notation also translates logical constants. `T`- and `F`-poles assigned to atomic propositions indicate the affirmation and negation, respectively, of the corresponding atomic propositions. Single `T`-pole-groups (`F`-pole-groups) list the non-redundant subconditions that constitute a sufficient condition for truth (falsehood). These single pole-groups, in turn, are arranged into lists of the sufficient conditions for truth or falsehood. Making use of the features of this pole-group notation, one obtains the following ideal diagram as a representation of the logical form of (1):

$$\begin{gathered}
\mathtt{T} - \left\{ \begin{array}{c} \{\mathtt{T} - P,\ \mathtt{F} - Q\}, \\ \{\mathtt{F} - P,\ \mathtt{T} - Q\}, \\ \left\{ \genfrac{}{}{0pt}{2}{\mathtt{T}-P,\ \mathtt{T}-R}{\mathtt{T}-Q,\ \mathtt{T}-R} \right\} \end{array} \right\} \\
\mathtt{F} - \left\{ \begin{array}{c} \{\mathtt{F} - P, F - Q\}, \\ \{\mathtt{T} - P, \mathtt{T} - Q, \mathtt{F} - R\} \end{array} \right\}
\end{gathered}$$

**Fig. 1.** Ideal diagram of (1)

From this diagram, it is possible to directly read off the conditions for the truth and falsehood of admissible instances of formula (1). The following is a paraphrase of this ideal diagram:[3]

---

[3] I abstain here from cumbersome references to instances of atomic formulas. Thus, I refer to $P$ instead of "an admissible instance of $P$", etc. I also abstain from specifying the trivial algorithm for paraphrasing ideal diagrams of propositional logic.

An admissible instance of formula (1) is true iff
  - $P$ is true and $Q$ is false, or
  - $P$ is false and $Q$ is true, or
  - one of the following alternatives:
    $P$ is true and $R$ is true/$Q$ is true and $R$ is true.

An admissible instance of formula (1) is false iff
  - $P$ is false and $Q$ is false, or
  - $P$ is true and $Q$ is true and $R$ is false.

This paraphrase of the ideal diagram is valid for all formulas equivalent to (1). Unlike the paraphrases of propositional formulas in general, the paraphrase of the ideal diagram of a propositional formula identifies common features of the models and counter-models for all formulas in the set of logically equivalent formulas. Instead of specifying *single* interpretations as models and counter-models, as is the case in model theory, the ideal diagram *describes* the properties of *sets* of models and counter-models. This difference is significant when there are an infinite number of models and counter-models, as in FOL.

In the case of logical theorems, Wittgenstein's `TF`-notation makes explicit that the conditions for truth and falsehood do not depend on the truth values of any atomic propositions. $P \vee \neg P$ and $Q \vee \neg Q \vee (R \wedge S)$, for example, are logically equivalent theorems. As long as one is interpreting logical formulas, the interpretation of the first formula seems to depend on the truth values of instances of $P$, whereas the interpretation of the second seems to depend on the truth values of instances of $Q$, $R$ and $S$. According to WIL, however, this is an illusion caused by the outer forms of the formulas. As soon as one is relating semantics not to initial formulas but rather to ideal diagrams, it becomes clear that logical theorems and their instances do not depend on the truth values of any atomic propositions. They all have the same conditions for truth and falsehood; they all say the same thing, namely, nothing. This becomes apparent upon the application of a reduction algorithm that deletes atomic propositions in the process of generating the ideal diagram. One may use $T - \{\square\}$ to represent that the conditions for truth comprise the entire space of logical possibilities, whereas $F - \{\blacksquare\}$ (or, alternatively, $F - \{\}$) may be used to represent that the conditions for falsehood are not included within the space of what is logically possible. This is the shared logical form of all "logical propositions" that Russell was unable to present within his symbolism (cf. p. 2).

By applying the well-known Quine-McCluskey reduction algorithm in propositional logic and several rather trivial rules for generating ideal diagrams within Wittgenstein's `TF`-notation, the concept of proof in the WIL version of propositional logic is fully defined. From this, it is clear what must be achieved within FOL: one must find a procedure for generating minimal DNFs in FOL (= FOLDNFs) and translate the resulting sets of minimal FOLDNFs into ideal diagrams in Wittgenstein's notation. [Lampert (2017b)] prescribes how to achieve this for the fragment of FOL that starts from formulas that do not contain any dyadic sentential connectives in the scope of quantifiers. [Lampert (2017c)] generalizes this prescription to a decision procedure for the FOL fragment that is translatable into disjunctions of conjunctions of formulas that do not contain $\vee$ in the scope of quantifiers. [Lampert (2017a)] defines an effective procedure for generating FOLDNFs in general and specifies an effective procedure for minimizing them.[4] However, this procedure does not fully satisfy the requirements for a proof procedure of WIL because it does not fully minimize the FOLDNFs in every case. The task of finding such a procedure remains an open problem. In the following section, I first define the syntactic properties of minimal FOLDNFs and then specify (i) how to translate them into ideal diagrams and (ii) how to paraphrase those ideal diagrams. This discussion should clarify the meaning of a representation of a logical form in FOL, although to date, no general algorithm has been specified that can generate such representations in all cases.

## 5 Ideal Diagrams II - First-Order Logic

Minimal FOLDNFs are defined in terms of primary formulas, which correspond to negated and non-negated atomic formulas in the DNFs of propositional logic. The term negation normal forms (NNFs) refers to formulas that contain $\neg$ only directly to the left of atomic propositional functions and $\wedge$ and $\vee$ only as dyadic connectives.

**Primary Formula:**

1. An NNF that does not contain $\wedge$ or $\vee$ is a primary formula.
2. NNFs that contain $\wedge$ or $\vee$ are primary formulas iff they satisfy the following conditions:
   (a) Any conjunction of $n$ conjuncts ($n > 1$) is preceded by a sequence of existential quantifiers of minimal length 1, and all $n$ conjuncts contain each variable of the existential quantifiers of that sequence.
   (b) Any disjunction of $n$ disjuncts ($n > 1$) is preceded by a sequence of universal quantifiers of minimal length 1, and all $n$ disjuncts contain each variable of the universal quantifiers of that sequence.
3. Only NNFs that satisfy condition 1 or 2 are primary formulas.

Primary formulas represent the limit to which quantifiers can be driven inwards by applying PN laws, i.e., the equivalence laws that are used to generate prenex normal forms if applied in the opposite direction. Cases 2(a) and 2(b) above are the only cases in which PN laws cannot be applied to drive quantifiers any farther inwards. As will be shown below, primary formulas can be translated into diagrams within Wittgenstein's notation that satisfy the conditions for ideal representations given that no conjunct or disjunct is redundant. [Lampert (2017a)], Sect. 2, specifies an effective algorithm for generating

---

[4] This procedure as well as others are implemented at and can be applied via the link given in footnote 1.

FOLDNFs (i.e., disjunctions of conjunctions of primary formulas).[5] Thus, there is no difficulty in establishing this part of the procedure for generating ideal diagrams for FOL.

Minimal FOLDNFs are defined as follows:

**Minimal FOLDNF:** A minimal FOLDNF is a disjunction of conjunctions of primary formulas that satisfies the following condition: If any number of conjuncts or disjuncts (whether they occur inside the scope of quantifiers, i.e., within the primary formulas, or outside the scope of quantifiers) is deleted, then the resulting formula is not equivalent to the initial one.

Thus, no conjunct or disjunct is redundant in the case of minimal FOLDNFs. Defining a general procedure for generating the set of *minimal* FOLDNFs from FOLDNFs is the problematic part of implementing Wittgenstein's concept of proof within FOL.

The crucial difference between FOLDNFs and the DNFs of propositional logic is the use of primary formulas. In order to clarify how they contribute to specifying the properties of models and counter-models, I will describe how they can be translated into ideal diagrams in Wittgenstein's notation and then illustrate how to paraphrase those ideal diagrams. Consider first an example of a minimal primary formula to motivate its translation into some other notation:

$$\exists y(\forall x(\neg Fxx \lor Hxy) \land Gy) \tag{6}$$

Suppose that this primary formula is part of a minimal FOLDNF that is equivalent to an initial formula (the details of which are unimportant here). (6) does not satisfy the standards of WIL in two respects: (i) it is equivalent to all formulas obtained by renaming the variables, and (ii) it contains $\land$ and $\lor$, although these signs do not contribute to specifying truth conditions in the same way that they do when they occur outside the scope of quantifiers within FOLDNFs. Because (i) is true, the particular type of each variable does not contribute to representing the properties of models (or counter-models). Instead, it is the relations between bound variables and the positions at which those variables occur in the atomic propositional functions that represent the properties of models. Because (ii) is true, $\land$ and $\lor$ cannot be paraphrased in the same way both inside and outside the scope of quantifiers in FOLDNFs. Within primary formulas, $\land$ does not separate non-redundant subconditions of a sufficient condition for truth, and $\lor$ does not separate sufficient conditions for truth. (i) highlights how conditions 1 and 2(a), as listed on p. 3, contribute to the ability of the outer forms of logical formulas to disguise their logical forms, whereas (ii) highlights the contribution of conditions 2(a) and 2(b).

These problems can be solved by applying the following algorithm to translate primary formulas into their corresponding ideal two-dimensional "primary diagrams" in a Wittgensteinian notation:

---

[5] FOLDNFs are far less complex than Hintikka's distribute normal forms of FOL; cf. [Lampert (2017a)] for details.

1. Translate the propositional functions. Replace variables with numbers to indicate positions, and denote affirmation by `T` and negation by `F`.
2. Symbolize the relations between the bound variables and their positions. Use forks to connect the numbers of the positions as follows:
   (a) Open forks, e.g., $\begin{smallmatrix}/\\ \backslash\end{smallmatrix}$, connect the numbers of positions connected by a disjunction and bound by a universal quantifier.
   (b) Closed forks, e.g., $<$, connect the numbers corresponding to all other positions of one and the same variable.

   When a bound variable occurs only once within only one propositional function, no fork is needed.

Proceeding from inside to outside, this algorithm results in the following translation of (6):

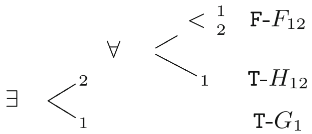


**Fig. 2.** Ideal diagram of (6)

Thus, the bound variables, $\vee$ and $\wedge$ are eliminated in favour of forks connecting the positions of the corresponding propositional functions.

This ideal diagram can now be paraphrased using a simple procedure that proceeds from outside to inside:

- Some object, the same in the second position of $H_{12}$ and in the first position of $G_1$, combined with all objects distributed among (i) the first and second positions of $F_{12}$ (where the same object appears in both positions) and (ii) the first position of $H_{12}$, makes the dyadic propositional function $F_{12}$ false, the dyadic propositional function $H_{12}$ true, and the monadic propositional function $G_1$ true.

This paraphrase clarifies how the properties of Fig. 2 identify the properties of models. Open forks represent distributions of objects, whereas closed forks indicate identical objects in different positions.

Here, I omit the cumbersome but trivial specification of (i) the general algorithm for translating FOLDNFs into ideal diagrams of FOL and (ii) the reading algorithm for ideal diagrams of FOL. It should be clear that the crucial problem encountered in an attempt to detail the proof procedure for WIL is specifying a procedure for fully minimizing FOLDNFs.

The translation of a minimal FOLDNF results in a finite ideal diagram in the `TF`-pole-group notation (cf. p. 5), in which each finite group describes the properties of a possibly infinite set of models (or counter-models) and each primary ideal diagram describes certain properties that all (counter-)models in such a set share. Instead of referring to an infinite number of models that may, in turn, involve infinite domains and infinite interpretations of atomic propositional functions, finite descriptions of the properties of such infinite sets are provided, independently of and prior to the evaluation of interpretations. The logical form of the models is identified, thus making it superfluous to explicitly refer to infinity. From the perspective of computability, this is crucial and desirable.

## 6 Application of Logic - The Question of Adequate Formalization

In this section, WIL is applied for the analysis of the logical forms of ordinary propositions. Two examples are presented that illustrate the misleading form thesis with respect to ordinary propositions, and the standard for adequate formalization according to WIL is explained. I assume the logic to be FOL. Therefore, the logical forms are restricted to logical forms that are expressible in terms of first-order formulas.

Example 1 concerns propositions of the following form:

All F's of G's are F's of H's. (7)

This form assumes that $F$, $G$ and $H$ are variables of atomic propositional functions. The following propositions are instances of (7):

All children of mothers are children of fathers. (8)

All heads of horses are heads of animals. (9)

All bets on winning numbers are bets on prime numbers. (10)

The *logical* forms of (8) to (10) must be independent of any specific internal relations between the meanings of the concepts invoked. From a logical standpoint, any possible combinatoric extension of these concepts is logically possible regardless of how strange, or even inconceivable, such a state of affairs would be. Therefore, mothers also being fathers, horses not being animals and headless horses are all *logical possibilities*. However, this does not mean that such strange possibilities correspond to *truth* conditions of their respective sentences. Instead, according to WIL, distinguishing between logical possibilities that make a sentence true and those that falsify it is a question of adequate logical formalization. The logical form shows how the truth conditions of a complex proposition depend on logically possible extensions of its atomic propositional functions.

The following two logical formulas are reasonable candidates for a logical formalization of propositions instantiating (7):[6]

---

[6] Standard logic textbooks, such as [Copi (1979)], p. 131f., or [Lemmon (1998)], p. 131f., formalize (9) by (11); by contrast, [Wengert (1974)] argues that only (8) should be formalized by (11), whereas (9) should be formalized by (12).

$$\forall x(\exists y(Fxy \land Gy) \rightarrow \exists z(Fxz \land Hz)) \tag{11}$$

$$\forall x \forall y((Fxy \land Gy) \rightarrow (Fxy \land Hy)) \tag{12}$$

WIL requires ideal diagrams to make explicit the conditions for the truth and falsehood of the formalized propositions in relation to certain atomic propositional functions. According to WIL, it is not logical formulas but ideal diagrams that are judged to be adequate or inadequate as representations of the logical forms of formalized propositions. Consequently, it is possible to assign *unique* logical forms to unambiguous propositions within this framework. There is no need for formalization criteria that call for a similarity between logical and grammatical forms to allow one to choose among logically equivalent formulas (cf., e.g., [Peregrin and Svoboda (2017)], p. 73).

For simplicity, I will write down only the ideal diagrams for the falsehood conditions of (11) and (12) (cf. Figs. 3 and 4, respectively). The corresponding representations of the truth conditions are symmetrical in this case.

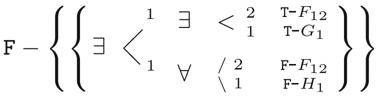


**Fig. 3.** $F$-pole-groups of the ideal diagram of (11)

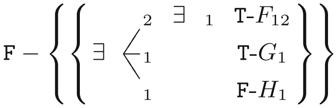


**Fig. 4.** $F$-pole-groups of the ideal diagram of (12)

According to an understanding of (8) in which "mother" and "father" refer to "biological mother" and "biological father", respectively, one can view Fig. 3 as an adequate formalization of the conditions for the falsehood of (8), whereas Fig. 4 can be seen as an adequate formalization of the conditions for the falsehood of (9). The difference is that (9) is false if there exists a head that is the head of a horse that is not an animal, whereas (8) is not false if there exists a child that is a child of a mother who is not a father. Instead, (8) is false only if some child exists who is a child of a mother but not of a father. In the case of (10), both Figs. 3 and 4 present reasonable forms for paraphrases of the conditions for falsehood. This situation shows that the meaning of (10) is ambiguous. Overall, this discussion demonstrates that the shared *outer form* (7) does not determine a unique logical form. This, however, does not mean that it is not reasonable to assign a logical form to a certain *ordinary proposition* with respect to a given set of atomic propositional functions. Instead, WIL provides the tools to do so while clarifying the conditions for the truth and falsehood of the initial propositions.

Example 1 has demonstrated that the outer form of a proposition does not determine a unique logical form. The following example illustrates that the outer form of a proposition also does not determine whether a proposition has a proper logical form *at all*. Thus, WIL provides the tools not only to express the conditions for the truth and falsehood of propositions within a logical framework but also to make explicit that certain propositions are not expressible within this framework.

Consider propositions of the following form:

$$\text{If someone (is in relation) F (to) a G, then a G exists.} \tag{13}$$

The following propositions are instances of (13) (cf. [Montague (1966)], p. 266, and [Quine (1960)], Sect. 30):

$$\text{If someone loves a women, then a woman exists.} \tag{14}$$

$$\text{If someone seeks a unicorn, then a unicorn exists.} \tag{15}$$

(13) can be translated into the following FOL formula:

$$\exists x \exists y (Fxy \wedge Gy \rightarrow Gy) \tag{16}$$

(16) is a logical theorem. Thus, it seems reasonable to formalize (14) by the ideal diagram of logical theoremhood (i.e., a diagram with empty conditions for falsehood). However, this is not the case for (15), which will most likely be judged to be false. This is why "x seeks y" is commonly regarded as an inadmissible instance of an atomic function variable within FOL.[7] From this it follows that logical forms cannot be assigned to propositions involving such a predicate. Therefore, (15) has no proper logical form, whereas (14) does.

The outer form of a proposition determines neither a unique logical form (cf. (7)) nor whether various propositions of a certain form share a logical form at all (cf. (13)). This is even true in cases of instances of provable formulas. The equivalence of the conditions for truth and falsehood is the criterion for adequate formalization when applying logic to propositions, and this cannot be judged without first generating ideal diagrams. According to WIL, interpretation comes last, not first.

## 7 Application of Logic - The Question of Decidability

A typical argument against Wittgenstein's conception of logic asserts that his understanding of logical proof assumes decidability, which is in conflict with

---

[7] According to [Quine (1960)], Sect. 30, a predicate such as "x seeks y" does not refer to a set of pairs and, thus, does not satisfy the principle of extensionality. However, the question is how one can know this without referring to some failure of logical formalization. For our purposes, it is sufficient to note that mere instantiation of logical formulas does not guarantee that those instances behave in accordance with the laws of logic. Therefore, one must distinguish between admissible and inadmissible instances. According to WIL, instances are inadmissible if they are not judged to be true despite instantiating provable formulas.

the Church-Turing theorem (cf., among others, [Landini (2007)], p. 118, and [Potter (2009)], p. 181f). However, this argument is not conclusive because the undecidability proof of FOL makes assumptions that WIL rejects.

Turing's undecidability proof relies on a formalization of the code of Turing machines. Furthermore, it relies on a claim that propositions about Turing machines that result from substituting propositional functions for function variables in provable formulas are true. To justify this claim, Turing explicitly refers to the following general principle (cf. [Turing (1936)], p. 262):

> If we substitute any propositional functions for function variables in a provable formula, we obtain a true proposition.

As argued in the previous section, this principle applies only to "admissible instances", i.e., instances that have a certain logical form and, hence, have conditions for truth and falsehood that are expressible within FOL. However, the expressibility within FOL is questionable in the case of diagonalization, which produces self-referential propositions. For example, it is common to reject "This proposition is not true" as an admissible instance of the function variable $P$ in the provable formula $P \leftrightarrow P$. Like other undecidability proofs, Turing's undecidability proof rests on diagonalization. Turing argues that a Turing machine $\mathcal{E}$ that decides on *any* logical formula cannot exist because the decision on a formula involving the formalization of $\mathcal{E}$ in the diagonal case cannot correspond to the behavior of certain machines involving $\mathcal{E}$. The quoted principle does not demonstrate that Turing's formalization of Turing machines is adequate in the diagonal case. Therefore, his inference of the non-existence of a Turing machine $\mathcal{E}$ is a fallacy.[8]

## 8 Wittgenstein and Peirce

There are many similarities and differences between WIL and PIL.[9] I refer only to the most essential ones in the following.

Peirce distinguished two purposes of logic: to investigate logical theories and to aid in the drawing of inferences. A logical calculus serves the latter purpose, whereas a logical system serves the former. Such a system should explain what is expressible through logic. To this end, it must not allow for 'any superfluity of symbols' ([Peirce (1931–1958)], 4.373):

---

[8] In fact, I have detailed a decision procedure for pure FOL without identity on the basis of a Wittgensteinian conception of proof (cf. the link given in footnote 1). For the details of a Wittgensteinian critique of undecidability proofs, cf. [Lampert (2017d)].

[9] Cf., in particular, [Shin (2002)] and [Dau (2006)] for detailed elaborations of PIL. [Pietarinen (2006)] provides a game-theoretic interpretation of PIL and relates this interpretation to the later work of Wittgenstein. By contrast, I refer to the early work of Wittgenstein and his conception of a logical proof as a mechanical transformation into ideal diagrams.

> It should be recognized as a defect of a system intended for logical study that it has two ways of expressing the same fact, or any superfluity of symbols, although it would not be a serious defect for a calculus to have two ways of expressing a fact.

Similar to Peirce's distinction between the calculi of symbolic logic and his existential graphs, Wittgenstein drew a distinction between the axiomatic proof method and his own proof method (cf. [Wittgenstein (1979)], p. 109, and [Wittgenstein (1994)], 6.125). On the one hand, he emphasized that the two methods are equivalent (i.e., do not differ in their results; cf. [Wittgenstein (1994)], 6.125f., and [Wittgenstein (1994)], p. 80). On the other hand, he regarded the traditional method of symbolization, which allows for 'a plurality' of equivalent symbols, as defective with regard to the analysis of propositions ([Wittgenstein (1997)], p. 102[3]; see also p. 93[1] and [Wittgenstein (1994)], 5.43):

> If $p = not - not - p$ etc., this shows that the traditional method of symbolism is wrong, since it allows a plurality of symbols with the same sense; and thence it follows that, in analyzing such propositions, we must not be guided by Russell's method of symbolizing.

Iconic logic can be distinguished from symbolic logic by the search for a procedure for transforming logical formulas into ideal diagrams that do not permit any 'plurality' or 'superfluity' of symbols.

Wittgenstein considered the need for a theory of deduction and for semantics as foundations of pure logic to be a result of a deficient symbolism. He desired to eliminate the need for semantic foundations by identifying 'the sense' of propositions (i.e., the conditions for their truth and falsehood) by means of iconic features of ideal diagrams. According to Wittgenstein, it is not reality (facts) but rather the *logical possibilities* of truth and falsehood that are represented by ideal diagrams. This is why WIL introduces bipolarity as a fundamental property of a proper logical notation, whereas Peirce claims that 'symmetry always involves superfluity' and that symmetries 'are great evils' for 'the purposes of analysis' (cf. [Peirce (1931–1958)], 4.375). In this respect, WIL differs from Peirce's existential graphs, which seem to instead be guided by the desire to represent reality (cf. the above quote from [Peirce (1931–1958)], 4.373, and [Shin (2002)], p. 52). It is for this reason that existential graphs do not correspond to FOLDNFs, nor even to NNFs (given an endoporeutic reading). The question of how to read or interpret existential graphs is a controversial one. From a Wittgensteinian perspective, this very controversy indicates that these graphs share some of the deficiencies of the conventional logical symbolism.

## References

Copi, I.: Symbolic Logic. Macmillan, New York (1979)

Dau, F.: Mathematical Logic with Diagrams - Based on the Existential Graphs of Peirce (2006). http://www.dr-dau.net/eg_readings.shtml

Etechemendy, J.: The Concept of Logical Conequence. CSLI Publications, Standford (1999)

Lampert, T.: Minimizing disjunctive normal forms of first-order logic. Log. J. IGPL **25**(3), 325–347 (2017)

Lampert, T.: Wittgenstein's *ab*-notation: an iconic proof procedure. Hist. Philos. Log. **38**(3), 239–262 (2017)

Lampert, T.: A decision procedure for Herbrand formulae without skolemization, pp. 1–30 (2017). https://arxiv.org/abs/1709.00191

Lampert, T.: Wittgenstein and Gödel: an attempt to make 'Wittgenstein's objection' reasonable. Philosophia Mathematica **25**(3), 1–22 (2017). https://doi.org/10.1093/philmat/nkx017

Landini, G.: Wittgenstein's Apprenticeship with Russell. Cambridge University Press, Cambridge (2007)

Lemmon, E.J.: Beginning Logic. Hackett, Indianapolis (1998)

Montague, R.: The proper treatment of quantification in ordinary English. In: Thomason, R.H. (ed.) Formal Philosophy. Selected Papers, pp. 247–270. Yale University Press, New Haven (1966)

Peirce, C.S.: Collected Papers of Charles Sanders Peirce. In: Hartshorne, C., Weiss, P. (eds) Harvard University Press, Cambridge (1931–1958)

Peregrin, J., Svoboda, V.: Reflective Equilibrium and the Principles of Logical Analysis. Routledge, New York (2017)

Pietarinen, A.-V.: Signs of Logic. Springer, Dordrecht (2006)

Potter, M.: Wittgenstein's Notes on Logic. Oxford University Press, Oxford (2009)

Quine, W.V.O.: Word and Object. MIT Press, Cambridge (1960)

Russell, B.: The Principles of Mathematics, 2nd edn from 1937. Routledge, London (1992)

Shin, S.J.: The Iconic Logic of Peirce's Graphs. The MIT Press, Cambridge (2002)

Turing, A.: On computable numbers, with an application to the Entscheidungsproblem. Proc. Lond. Math. Soc. **2**(42), 230–265 (1936)

Wengert, R.G.: Schematizing de Morgan's argument. Notre Dame J. Form. Log. **1**, 165–166 (1974)

Wittgenstein, L.: Cambridge Letters. Blackwell, Oxford (1997)

Wittgenstein, L.: Notebooks 1914-1916. Blackwell, Oxford (1979)

Wittgenstein, L.: Tractatus Logico-Philosophicus. Routledge, London (1994)

Wittgenstein, L.: Wittgenstein and the Vienna Circle. Blackwell, Basil (1979)

# Aristotelian and Duality Relations Beyond the Square of Opposition

Lorenz Demey[1(✉)] and Hans Smessaert[2]

[1] Center for Logic and Analytic Philosophy, KU Leuven, Leuven, Belgium
`lorenz.demey@kuleuven.be`
[2] Department of Linguistics, KU Leuven, Leuven, Belgium
`hans.smessaert@kuleuven.be`

**Abstract.** Nearly all squares of opposition found in the literature represent both the Aristotelian relations and the duality relations, and exhibit a very close correspondence between both types of logical relations. This paper investigates the interplay between Aristotelian and duality relations in diagrams beyond the square. In particular, we study a Buridan octagon, a Lenzen octagon, a Keynes-Johnson octagon and a Moretti octagon. Each of these octagons is a natural extension of the square, both from an Aristotelian perspective and from a duality perspective. The results of our comparative analysis turn out to be highly nuanced.

**Keywords:** Aristotelian relations · Duality relations
Square of opposition · Aristotelian diagram · Duality diagram
Logical geometry

## 1 Introduction

The square of opposition represents four propositions, and certain logical relations holding between them. This diagram has a long and well-documented history in philosophy and logic [36]. In contemporary (analytic) philosophy, it has been used in various areas, such as philosophy of language, epistemology, philosophy of religion, ethics, and philosophy of law. In logic, the square of opposition has been used to study systems of modal logic, various non-classical logics, probabilistic and fuzzy logics, and logics of rational agency. Finally, because of the ubiquity of the logical relations that it represents, the square is nowadays also frequently used outside the boundaries of philosophy and logic, in disciplines such as psychology, linguistics and computer science. A comprehensive overview of this wide diversity of applications (including many bibliographic references) can be found in [15,16]. The square of opposition visually represents the *Aristotelian relations*: contradiction, contrariety, subcontrariety, and subalternation. However, most—nearly all—squares that appear in the literature also exhibit another type of logical relations, viz. the *duality relations*: internal negation, external negation and duality. Based on the concrete diagrams found in the literature, the notions of *Aristotelian square* and *duality square* thus seem to be

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 640–656, 2018.
https://doi.org/10.1007/978-3-319-91376-6_57

almost co-extensional with each other. Nevertheless, there also seem to be clear conceptual differences between both types of logical diagrams.

The research program of *logical geometry* is concerned with the systematic study of logical diagrams in general, and Aristotelian diagrams and duality diagrams in particular. We investigate these diagrams using cognitive and geometric notions, such as informational vs. computational equivalence [12,14], Euclidean distance [16,44], vertex-first projections [10] and subdiagrams [6,42]. On the logical side, we focus on issues such as diagram informativity [43], logic-sensitivity [8], diagram classification [45] and Boolean structure [15,46]. The visual and logical properties of Aristotelian diagrams and duality diagrams are thus relatively well-understood in isolation. However, we do not yet have a clear picture of the precise interconnections between these two types of logical diagrams. Smessaert [41] has achieved some promising results in this direction, by moving beyond the square of opposition and focusing on a specific hexagon of opposition.

The main goal of this paper is to further advance this line of research, by analyzing the interplay between Aristotelian and duality relations in several octagons of opposition. These octagons will be shown to be very natural extensions/generalizations of the classical square of opposition, both from an Aristotelian perspective and from a duality perspective. With respect to the latter, we will discuss two main generalizations of 'classical' duality, viz. composed operator duality and generalized Post duality. This approach constitutes a major improvement over that of [41], since the comparative analysis there is based on a hexagon, which naturally extends the classical square from an Aristotelian perspective, but arguably *not* from a duality perspective. Consequently, the comparative analysis in this paper will provide a more solid basis for drawing conclusions regarding the interconnections between Aristotelian and duality diagrams. In particular, we will first focus on the individual Aristotelian and duality relations, and argue that the systematic correspondence, as found in the square of opposition, is lost in the octagons (albeit to varying degrees).[1] Furthermore, there is no systematic correspondence at the level of entire diagrams either. Nevertheless, we show that at a higher level of abstraction, the correspondence does seem to remain intact (again, to varying degrees).

The paper is organized as follows. Section 2 describes the interplay between Aristotelian and duality relations in the classical square of opposition. Next, Sect. 3 discusses the (in)dependence of these two types of relations, and examines Smessaert's [41] comparative analysis based on a hexagon of opposition. Sections 4 and 5 constitute the core of this paper. Section 4 is concerned with composed operator duality, and analyzes the interplay between Aristotelian and duality relations in a Buridan octagon and a Lenzen octagon. Next, Sect. 5 focuses on generalized Post duality, and analyzes the interplay between Aristotelian and duality relations in a Keynes-Johnson octagon and a Moretti octagon. Finally,

---

[1] Similar conclusions were reached in [41], but there, one could still object that the loss of correspondence beyond the square is merely due to the fact that from a duality perspective, the hexagon is not a natural generalization of the square of opposition. Such an objection cannot be raised against the conclusions drawn in this paper.

Sect. 6 summarizes the results obtained in this paper, and mentions some questions for future research.

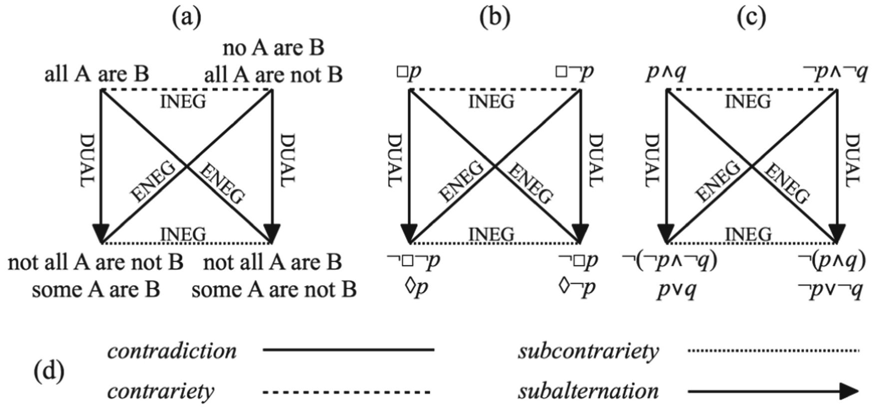


**Fig. 1.** Squares of opposition for (a) syllogistics, (b) modal logic, (c) propositional logic; (d) code for visualizing the Aristotelian relations.

## 2 Aristotelian and Duality Squares

We begin by considering the three most well-known squares of opposition. Without a doubt, the oldest and most frequently used square of opposition is that for *syllogistics*, as shown in Fig. 1(a). Both with respect to history and to frequency of use, a close second is the square of opposition for *modal logic*, as shown in Fig. 1(b). Furthermore, with the seminal work of authors such as Boole, De Morgan and Frege in the 19th and early 20th century also came the square of opposition for *propositional logic*, as shown in Fig. 1(c). Each of these square diagrams exhibits four key propositions of their underlying logical system, and the Aristotelian relations holding between them. These relations can be defined on various levels of generality and abstractness [13], but for our current purposes it will suffice to consider the most informal definition: two propositions are

| | | | |
|---|---|---|---|
| *contradictory* | iff | they cannot be true together<br>they cannot be false together, | and |
| *contrary* | iff | they cannot be true together<br>they can be false together, | and |
| *subcontrary* | iff | they can be true together<br>they cannot be false together, | and |
| *in subalternation* | iff | the first one entails the second one<br>the second one does not entail the first one. | and |

These relations will be abbreviated as $CD$, $C$, $SC$ and $SA$, respectively, and visualized according to the code in Fig. 1(d). For example, in Fig. 1 we observe that $CD($*some A are B*, *no A are B*$)$ in the syllogistic square, $C(\Box p, \Box\neg p)$ and $SC(\Diamond p, \Diamond\neg p)$ in the modal square, and $SA(p \wedge q, p \vee q)$ in the propositional square.

The contradiction relation is the most important Aristotelian relation,[2] and accordingly, it plays a crucial role in Aristotelian diagrams. Each proposition $\varphi$ has a unique contradictory (up to logical equivalence), viz. $\neg\varphi$. The square of opposition, and almost all other Aristotelian diagrams found in the literature as well, are closed under contradiction: if the diagram contains $\varphi$, then it also contains $\neg\varphi$.[3] The propositions occurring in an Aristotelian diagram can thus naturally be grouped into *pairs of contradictory propositions* (PCDs). Consequently, a square of opposition should not simply be seen as consisting of 4 'individual' propositions, but rather of 2 PCDs. This perspective also suggests a natural way of extending the square, viz. by adding more PCDs. We thus go from 2 PCDs to 3 PCDS, 4 PCDs, etc.—or in more geometric/diagrammatic terms: from square to hexagon, octagon, etc.[4]

The squares of opposition in Fig. 1(a–c) not only represent the Aristotelian relations, but also the duality relations. Just as before, these relations can be defined on various levels of generality and abstractness [13], but for our current purposes it will again suffice to consider the most informal definition. Suppose that $\varphi$ and $\psi$ are the results of applying $n$-ary operators $O_\varphi$ and $O_\psi$ to the same $n$ propositions $\alpha_1, \ldots, \alpha_n$, i.e. $\varphi \equiv O_\varphi(\alpha_1, \ldots, \alpha_n)$ and $\psi \equiv O_\psi(\alpha_1, \ldots, \alpha_n)$. We then say that $\varphi$ and $\psi$ are each other's

$$\begin{array}{lllll} \textit{external negation} & \text{iff} & O_\varphi(\alpha_1, \ldots, \alpha_n) & \equiv & \neg O_\psi(\ \alpha_1, \ldots,\ \alpha_n), \\ \textit{internal negation} & \text{iff} & O_\varphi(\alpha_1, \ldots, \alpha_n) & \equiv & \ \ O_\psi(\neg\alpha_1, \ldots, \neg\alpha_n), \\ \textit{dual} & \text{iff} & O_\varphi(\alpha_1, \ldots, \alpha_n) & \equiv & \neg O_\psi(\neg\alpha_1, \ldots, \neg\alpha_n). \end{array}$$

These relations will be abbreviated as ENEG, INEG and DUAL, respectively. Note that INEG operates on *all* propositions $\alpha_1, \ldots, \alpha_n$. In Fig. 1 we see that ENEG(*some A are B*, *no A are B*) in the syllogistic square, INEG$(\Box p, \Box\neg p)$ and INEG$(\Diamond p, \Diamond\neg p)$ in the modal square, and DUAL$(p \wedge q, p \vee q)$ in the propositional square.

The logical behavior of the duality relations is well-understood [11,45]. In particular, these relations are all functional (up to logical equivalence); for example, if INEG$(\varphi, \psi_1)$ and INEG$(\varphi, \psi_2)$, then $\psi_1 \equiv \psi_2$. Hence we can also view them as functions, and write, for example, $\psi =$ INEG$(\varphi)$ instead of INEG$(\varphi, \psi)$. Furthermore, since the INEG-relation is symmetrical—i.e. INEG$(\varphi, \psi)$ iff INEG$(\psi, \varphi)$—,

---

[2] Note that the definitions of contrariety and subcontrariety can both be seen as weakened versions of that of contradiction. It can also be shown that contradiction is the most informative of the Aristotelian relations [43].

[3] Furthermore, the contradiction relation is usually visualized by means of *central symmetry*, so that all pairs of contradictory propositions are represented by diagonals that intersect each other in the Aristotelian diagram's center of symmetry [10,12,14].

[4] In this paper we will not distinguish between different geometrical representations of the same set of PCDs. For example: (i) 3 PCDs can be visualized as a hexagon or as an octahedron; (ii) 4 PCDs can be visualized as an octagon or as a cube [12,14].

the INEG-function is idempotent: INEG(INEG($\varphi$)) = $\varphi$. (All of this applies not only to INEG, but also to ENEG and DUAL.) In sum, the three duality functions, together with the identity function ID (defined by ID($\varphi$) := $\varphi$ for all $\varphi$) form a Klein 4-group under composition ($\circ$) [1,37], with the following Cayley table:

| $\circ$ | ID | ENEG | INEG | DUAL |
|---|---|---|---|---|
| ID | ID | ENEG | INEG | DUAL |
| ENEG | ENEG | ID | DUAL | INEG |
| INEG | INEG | DUAL | ID | ENEG |
| DUAL | DUAL | INEG | ENEG | ID |

The Klein 4-group is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. This group-theoretical isomorphism gives us a firm syntactic handle on the duality relations: each copy of $\mathbb{Z}_2$ governs an independent negation position: the first copy corresponds to external negation, and the second corresponds to internal negation.[5] This also suggests a natural way of extending duality behavior beyond the square of opposition (i.e. beyond the Klein 4-group), viz. by adding more independent negation positions (i.e. by adding more copies of $\mathbb{Z}_2$). We thus go from $\mathbb{Z}_2 \times \mathbb{Z}_2$ (2 negation positions, yielding a group of $2^2 = 4$ duality functions) to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ (3 negation positions, yielding a group of $2^3 = 8$ duality functions), etc.

If we now bring the Aristotelian and duality perspectives together, we see that the squares of opposition in Fig. 1(a–c) exhibit a highly uniform correspondence between both types of logical relations. In particular, there is a correspondence between (i) the Aristotelian relation *CD* and the duality relation ENEG, (ii) the Aristotelian relations *C* and *SC* and the duality relation INEG, and (iii) the Aristotelian relation *SA* and the duality relation DUAL. Each square thus gives rise to an Aristotelian/duality multigraph (ADM) as shown in Fig. 2. This ADM visualizes, for each combination of an Aristotelian and a duality relation, how many times that specific combination occurs in the square of opposition.[6] Although the correspondence between Aristotelian and duality relations is not perfect, the ADM clearly shows that it is still highly regular. Using graph-theoretical terminology [18], the ADM for the square of opposition has 4 *connected components*, viz. $\{CD, \text{ENEG}\}$, $\{C, SC, \text{INEG}\}$, $\{SA, \text{DUAL}\}$ and $\{EQ, \text{ID}\}$. More concretely, each Aristotelian relation corresponds to a unique duality relation; vice

---

[5] Under the group-theoretical isomorphism between the Klein 4-group for the duality functions and $\mathbb{Z}_2 \times \mathbb{Z}_2$, ID corresponds to $(0,0)$ (apply no negations at all), ENEG to $(1,0)$ (only apply external negation), INEG to $(0,1)$ (only apply internal negation) and DUAL to $(1,1)$ (apply both external and internal negation). If the duality function $f$ corresponds to $(i,j) \in \mathbb{Z}_2 \times \mathbb{Z}_2$, we thus get $f(O(\alpha_1, \ldots, \alpha_n)) = \neg^i O(\neg^j \alpha_1, \ldots, \neg^j \alpha_n)$ (with the usual definitions $\neg^0 \varphi := \varphi$ and $\neg^1 \varphi := \neg\varphi$).

[6] Note that the ADM includes *EQ* (logical equivalence) as the Aristotelian counterpart of ID. Strictly speaking, *EQ* is not one of the Aristotelian relations, but it is closely related to them [43], and it is implicitly present whenever we write multiple, logically equivalent propositions in a single vertex of an Aristotelian diagram. (Each vertex thus has an *EQ*-loop to itself.) Note, in this context, that the square of opposition is sometimes also referred to as 'the square of opposition *and equipollence*' [33].

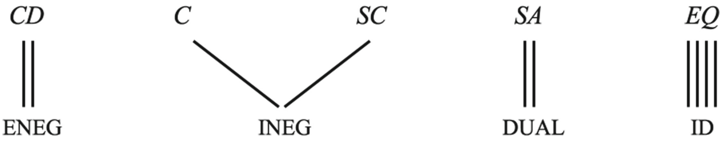


**Fig. 2.** Aristotelian/duality multigraph (ADM) for the classical square of opposition.

versa, the duality relations ENEG, DUAL and ID correspond to a unique Aristotelian relation, while INEG corresponds to two Aristotelian relations.

In summary, the well-known squares of opposition from Fig. 1(a–c) show that there is a clear correspondence between Aristotelian and duality considerations. At the level of *diagrams*, these squares of opposition are simultaneously Aristotelian squares and duality squares. At the level of the individual *relations*, the correspondence is summarized by the ADM in Fig. 2. Finally, it bears emphasizing that this correspondence can also be observed in more recent (and thus lesser-known) squares of opposition, such as those for public announcement logic [7], future contingents [21], definite descriptions [9], and rough set theory [48].

## 3 (In)dependence of Aristotelian and Duality Diagrams

Because of this correspondence, several authors [5,39,48] come close to outright identifying the two types of squares—e.g. by using Aristotelian terminology to describe the duality square (or vice versa), or by viewing one as a generalization of the other. The correspondence was already noted in medieval logic: influential authors such as Peter of Spain [4], William of Sherwood [29] and John Wyclif [19] discussed the mnemonic rhyme *pre contradic, post contra, pre postque subalter*, in which external negation (*pre*) is associated with contradiction, internal negation (*post*) with contrariety, and duality (*pre postque*) with subalternation.[7]

Despite this close correspondence, there are still some crucial differences between Aristotelian and duality diagrams [3,47]. Regarding the individual relations, it should be pointed out that (i) the duality relations are all symmetric, whereas the Aristotelian relation $SA$ is asymmetric, and that (ii) the duality relations are all functional, whereas the Aristotelian relations $C$, $SC$ and $SA$ are not (i.e. a single proposition can have multiple, non-equivalent contraries, subcontraries, and subalterns). Furthermore, the Aristotelian relations are far more

---

[7] This rhyme is incomplete, because as we have seen above (Fig. 2), internal negation (*post*) should not just be associated with contrariety, but also with subcontrariety [29, Footnote 54]. However, this omission can be explained in terms of the famous non-lexicalization of the O-corner [22]. The fact that *no A are B* is the internal negation of *all A are B* (i.e. *no* $\equiv$ *all* $\neg$, or in Latin: *nullus* $\equiv$ *omnis* $\neg$) is a contingent, empirical fact about English (resp. Latin), and should thus be captured by the rhyme. By contrast, the fact that *some A are not B* is the internal negation of *some A are B* (i.e. *some not* $\equiv$ *some* $\neg$, or in Latin: *aliquis non* $\equiv$ *aliquis* $\neg$) is almost analytically true, and thus need not be captured by the mnemotechnic rhyme.

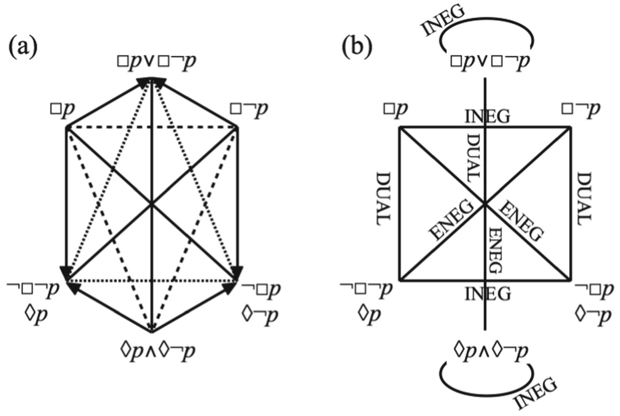


**Fig. 3.** (a) Aristotelian relations and (b) duality relations in the modal JSB hexagon.

sensitive to the details of the underlying logical system than the duality relations [8]. Consider, for example, the propositions $\Box p$ and $\Box \neg p$. In the normal modal logic $\mathsf{KD}$, these two propositions are contrary to each other, but in the weakest normal modal logic, $\mathsf{K}$, they do not stand in any Aristotelian relation at all [23]. Nevertheless, in both $\mathsf{KD}$ and $\mathsf{K}$, these two propositions are each other's internal negation. In general, as long as two logical systems have classical Boolean connectives, they will yield the same duality relations, even though they might yield vast differences in the Aristotelian relations.

Perhaps the most powerful way to argue for the independence of Aristotelian and duality diagrams consists in analyzing diagrams *beyond* the square. For example, Smessaert [41] has studied the interplay between Aristotelian and duality relations in a *hexagon* of opposition, as shown in Fig. 3.[8] From an Aristotelian perspective, this hexagon is a very natural extension of the square: it is obtained by adding one pair of contradictory propositions (PCD), thus moving from a diagram with 2 PCDs to one with 3 PCDs. This type of hexagon is very well-known [15,42]; it was first studied in the 1950s by Jacoby [25], Sesmat [40] and Blanché [2], and is therefore called a 'Jacoby-Sesmat-Blanché (JSB) hexagon'.

This hexagon clearly illustrates the discrepancy in functionality between the Aristotelian and the duality relations. For example, $\Box p$ has a unique internal negation, viz. $\Box \neg p$, but it has multiple contraries, e.g. $\Box \neg p$ and $\Diamond p \wedge \Diamond \neg p$. The contrariety between $\Box p$ and $\Box \neg p$ thus corresponds to an INEG-relation (just like in the square), but the contrariety between $\Box p$ and $\Diamond p \wedge \Diamond \neg p$ does not correspond to any duality relation at all (which we will denote as $\varnothing$). Similarly, $\Box p$ has a unique dual, viz. $\Diamond p$, but it has multiple subalterns, e.g. $\Diamond p$ and $\Box p \vee \Box \neg p$. The subalternation from $\Box p$ to $\Diamond p$ thus corresponds to a DUAL-relation (just like in

---

[8] For reasons of space, we only consider the *modal* hexagon, which extends the modal square in Fig. 1(b). In exactly the same way, one could also extend the other two squares in Fig. 1(a)/(c) to hexagons, and draw the same conclusions about them.

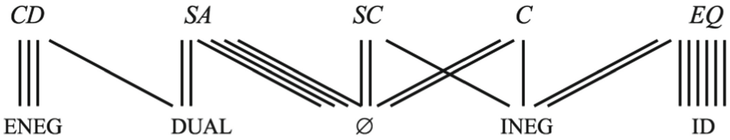


**Fig. 4.** Aristotelian/duality multigraph (ADM) for the modal JSB hexagon.

the square), but the subalternation between $\Box p$ to $\Box p \vee \Box\neg p$ does not correspond to any duality relation at all ($\varnothing$). Furthermore, $\Diamond p \wedge \Diamond\neg p$ turns out to be its own internal negation, so this INEG-relation will correspond to a logical equivalence ($EQ$). Consequently, $\Box p \vee \Box\neg p$ is not only the external negation of $\Diamond p \wedge \Diamond\neg p$, but also its dual; these ENEG- and DUAL-relations thus both correspond to the Aristotelian $CD$-relation. The entire configuration of Aristotelian and duality relations in the modal JSB hexagon is summarized by the ADM in Fig. 4.

By comparing the ADM for the square (Fig. 2) with that for the JSB hexagon (Fig. 4), we immediately see that the latter is much more 'cluttered'. Instead of having 4 connected components, the entire multigraph in Fig. 4 is connected. Each Aristotelian relation corresponds to multiple duality relations (or the complete absence of any duality relation, $\varnothing$); vice versa, DUAL corresponds to two Aristotelian relations, while INEG corresponds to two Aristotelian relations and logical equivalence ($EQ$). In sum: the systematic correspondence between Aristotelian and duality relations is completely lost in the JSB hexagon.

One might object to the conclusion of this analysis. After all, the JSB hexagon is a natural extension of the square from an Aristotelian perspective, but *not* from a duality perspective. The hexagon is obtained by adding an extra PCD, but in terms of duality, this does not correspond to adding an extra negation position. Consequently, the hexagon cannot be seen as a single, 'unified' duality diagram, but should rather be seen as the superposition of two separate, independent duality diagrams, viz. the original duality square and the extra PCD (which are classified in [45] as a CLCL1 duality square and a collapsed, self-internal duality square, respectively). The independence of these two duality diagrams is illustrated by the high number of edges involving $\varnothing$ in the ADM in Fig. 4.

In the remainder of the paper, we will thus consider diagrams that are natural extensions of the square of opposition from both an Aristotelian and a duality perspective. In particular, we focus on *octagons of opposition*: these extend the Aristotelian square (from 2 PCDs to 4 PCDs, i.e. from $2 \times 2 = 4$ to $4 \times 2 = 8$ propositions) as well as the duality square (from 2 negation positions to 3 negation positions, i.e. from $2^2 = 4$ to $2^3 = 8$ propositions). We thus consider a new duality group $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, which has been studied purely abstractly [35], but which also has two distinct concrete interpretations, viz. *composed operator duality* and *generalized Post duality*. These two types of duality, and the octagons that they give rise to, will be discussed in Sects. 4 and 5, respectively.

# 4 Octagons for Composed Operator Duality

Suppose that $\varphi$ is the result of applying an $n$-ary composed operator $O_1 \circ O_2$ to $n$ propositions $\alpha_1, \ldots, \alpha_n$, i.e. $\varphi \equiv (O_1 \circ O_2)(\alpha_1, \ldots, \alpha_n) = O_1(O_2(\alpha_1, \ldots, \alpha_n))$.[9] For example, in modal logic we can view $\Box(p \wedge q)$ as the result of applying the composed operator $\Box \circ \wedge$ to the propositions $p$ and $q$. (Westerståhl [47] discusses a linguistic example, viz. possessives with quantifiers; e.g. *some athletes of each country.*) By moving to composed operators, we have added an extra negation position, viz. *intermediate negation.* The proposition $O_1(O_2(\alpha_1, \ldots, \alpha_n))$ has

- a unique external negation (ENEG): $\neg O_1( \ O_2( \ \alpha_1, \ldots, \ \alpha_n))$,
- a unique intermediate negation (MNEG): $O_1(\neg O_2( \ \alpha_1, \ldots, \ \alpha_n))$,
- a unique internal negation (INEG): $O_1( \ O_2(\neg \alpha_1, \ldots, \neg \alpha_n))$.

Since each of these 3 independent negation positions may or may not be occupied, $O_1 \circ O_2$ gives rise to $2^3 = 8$ propositions in total, which exhibit a much richer duality behavior [6]. We now have three negation operations, and thus three pairwise combinations: ENEG∘INEG, ENEG∘MNEG, and MNEG∘INEG (abbreviated as EI, EM and MI, respectively). Finally, there is ENEG∘MNEG∘INEG (abbreviated as EMI), which operates on all three negation positions simultaneously.

## 4.1 The Buridan Octagon in Modal Syllogistics

In the logical works of the medieval philosopher John Buridan, we find three distinct octagons that exhibit composed operator duality [17,28,38]. We will focus on (a simplified version of) Buridan's modal octagon, which contains quantified *de re* modal propositions such as $\forall x \Box Px$. This proposition is the result of applying the composed operator $\forall \circ \Box$ to $Px$. This octagon can be thought of as capturing the interaction between the syllogistic square and the modal square from Fig. 1(a–b), and is thus a natural extension of both of these squares [17]. The logical behavior of this type of diagrams is well-studied; within the classification of Aristotelian diagrams, it is called a 'Buridan octagon' (for obvious historical reasons) [15].

The modal octagon is shown in Fig. 5(a).[10] For example, we observe that $\forall x \Box Px$ is contrary to three propositions, viz. (i) $\forall x \Box \neg Px$, (ii) $\forall x \neg \Box Px$, and (iii) $\neg \forall x \neg \Box \neg Px$. The first of these contrarieties corresponds to an INEG-relation, the second one to an MNEG-relation, and the third one to an EMI-relation. There are also four pairs of propositions that do not stand in any Aristotelian relation at all; Buridan himself called these *disparatae*; today, such pairs are called *unconnected* (*Un*) [43]. Two *Un*-pairs correspond to EI-relations, while the two others correspond to INEG-relations. The entire distribution of Aristotelian/duality relations in Buridan's modal octagon is summarized by the ADM in Fig. 6.

---

[9] If $O_2$ is $n$-ary, the composed operator $O_1 \circ O_2$ will also be $n$-ary. Furthermore, $O_1$ will be assumed to be unary, but this assumption is not essential.

[10] To avoid cluttering the diagrams, we will henceforth not explicitly show the *CD*- and ENEG-relations. These occur exactly at the diagram's diagonals, which intersect each other in the diagram's center of symmetry (recall Footnote 3).

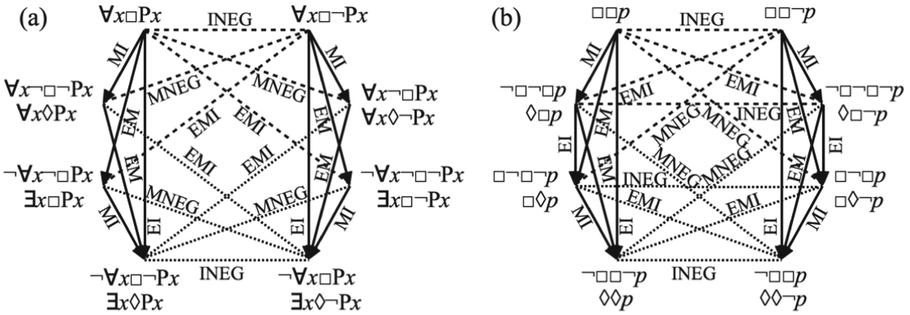


**Fig. 5.** (a) Buridan octagon in modal syllogistics; (b) Lenzen octagon in S4.2.

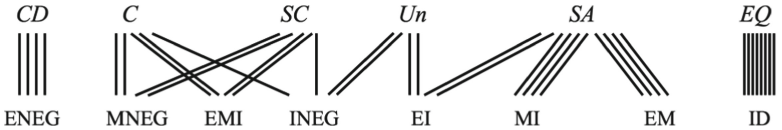


**Fig. 6.** ADM for the Buridan octagon in modal syllogistics.

This ADM has only 3 connected components. Two of these components are $\{CD, \text{ENEG}\}$ and $\{EQ, \text{ID}\}$, which represent two clear-cut correspondences between Aristotelian and duality relations. However, in all other cases, the correspondence is highly irregular. Apart from $CD$, all Aristotelian relations correspond to multiple duality relations. Vice versa, MI and EM correspond to a unique Aristotelian relation (just like ENEG), but all remaining duality relations correspond to multiple Aristotelian relations. All of this illustrates the lack of any systematic correspondence between Aristotelian and duality relations in Buridan's modal octagon. Furthermore, it should also be emphasized that this lack of correspondence cannot be due to Buridan's octagon purportedly not being a natural extension of the square of opposition from a duality perspective (unlike the JSB hexagon that was analyzed in Sect. 3). After all, we have already seen above that this octagon is a natural extension of the square from both an Aristotelian and a duality perspective.[11]

## 4.2 The Lenzen Octagon in S4.2

Another example of composed operator duality can be observed in the octagon in Fig. 5(b). This octagon is a natural extension of the modal square in Fig. 1(b): it is based on 'doubly modalized' propositions such as $\Box\Box p$, which can be seen as the result of applying the composed operator $\Box \circ \Box$ to the proposition $p$. In the well-known normal modal logic S4.2, these propositions stand in the Aristotelian relations shown in Fig. 5(b). (The key axiom of S4.2 is $\Diamond\Box p \rightarrow \Box\Diamond p$ [23].)

---

[11] Compare the ADMs for the modal JSB hexagon and Buridan's modal octagon in Figs. 4 and 6, and note the absence of $\varnothing$ in the latter.

Furthermore, some of these propositions can be simplified to 'singly modalized' propositions (e.g. $\Box\Box p$ is logically equivalent to $\Box p$ in S4.2), but we have not done so, in order to emphasize the composed operator duality exhibited by this octagon. This octagon belongs to a well-known type of Aristotelian diagrams, viz. the 'Lenzen octagons' (which is so-called because a diagram of this type was first used by Lenzen [30]). A Lenzen octagon has recently also been used in [9].

Looking at the octagon in Fig. 5(b), we observe, for example, that $\neg\Box\Box p$ is subcontrary to three propositions, viz. (i) $\neg\Box\Box\neg p$, (ii) $\Box\neg\Box\neg p$, and (iii) $\neg\Box\neg\Box p$. The first of these subcontrarieties corresponds to an INEG-relation, the second one to an EMI-relation, and the third one to an MNEG-relation. We also note that EMI$(\Box\Box p, \neg\Box\neg\Box\neg p)$ and EMI$(\Box\neg\Box\neg p, \neg\Box\Box p)$; the first of these EMI-relations corresponds to a contrariety, while the second one corresponds to a subcontrariety. The entire distribution of Aristotelian and duality relations in the Lenzen octagon in S4.2 is summarized by the ADM in Fig. 7.

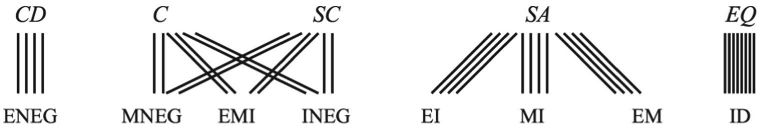


**Fig. 7.** ADM for the Lenzen octagon in S4.2.

This ADM shows that the correspondence between Aristotelian and duality relations in the Lenzen octagon in S4.2 is again quite irregular, although not as bad as in Buridan's modal octagon (recall Fig. 6). Apart from *CD*, all Aristotelian relations again correspond to multiple duality relations. Vice versa, however, only MNEG, EMI and INEG correspond to multiple Aristotelian relations—all other duality relations correspond to a unique Aristotelian relation.

When we compare the ADM for the Lenzen octagon (cf. Fig. 7) with that for the square of opposition (cf. Fig. 2), the similarities between both ADMs seem to prevail, rather than the dissimilarities. Both ADMs have 4 connected components, two of which are $\{CD, \text{ENEG}\}$ and $\{EQ, \text{ID}\}$, which represent clear-cut correspondences between Aristotelian and duality relations. Furthermore, the component $\{C, SC, \text{INEG}\}$ from the square has expanded into $\{C, SC, \text{MNEG}, \text{EMI}, \text{INEG}\}$. The composed operator duality relations MNEG, EMI and INEG thus jointly fulfill the role of the original INEG-relation, in corresponding to *C* and *SC*. Similarly, the component $\{SA, \text{DUAL}\}$ from the square has expanded into $\{SA, \text{MI}, \text{EM}, \text{EI}\}$, i.e. the composed operator duality relations MI, EM and EI jointly fulfill the role of the original DUAL-relation, in corresponding to *SA*.

## 5 Octagons for Generalized Post Duality

Recall that with classical duality, we assume that internal negation is applied to *all* argument positions, i.e. if $O$ is an $n$-ary operator, the internal negation of

$O(\alpha_1, \ldots, \alpha_n)$ is defined as $O(\neg\alpha_1, \ldots, \neg\alpha_n)$ (also cf. Footnote 5). However, we can also drop this assumption, and let internal negation apply to each argument position independently [24,31]. In the case of a binary operator $O$, we thus have 3 independent negation positions in total:[12] the proposition $O(\alpha_1, \alpha_2)$ has

- a unique external negation (ENEG): $\neg O(\ \alpha_1,\ \ \alpha_2)$,
- a unique first internal negation (INEG1): $O(\neg\alpha_1,\ \ \alpha_2)$,
- a unique second internal negation (INEG2): $O(\ \alpha_1, \neg\alpha_2)$.

Since each of these 3 independent negation positions may or may not be occupied, we obtain $2^3 = 8$ propositions in total, which again exhibit a much richer duality behavior. We now have three negation operations, and thus three pairwise combinations: ENEG ∘ INEG1, ENEG ∘ INEG2, and INEG1 ∘ INEG2 (abbreviated as EI1, EI2 and I12, respectively). Finally, there is ENEG ∘ INEG1 ∘ INEG2 (abbreviated as EI12), which operates on all three negation positions simultaneously.

## 5.1 The Keynes-Johnson Octagon in Syllogistics with Subject Negation

Classically, a categorical statement of the form *all A are B* is seen as the result of applying the unary operator *all A* to the predicate *B*—which gives rise to the square of opposition in Fig. 1(a). However, we can also view such a statement as the result of applying the binary operator *all* to the predicates *A* and *B*. If these two predicates can be negated independently, we obtain 8 propositions in total. Assuming that the extensions of *A* and *B* are neither empty nor the entire universe of discourse, these 8 propositions constitute the octagon of opposition shown in Fig. 8(a), which was first studied by Keynes [27] and Johnson [26]. The logical behavior of this Aristotelian diagram is well-studied [8,20]; from a classificatory perspective, it is called a 'Keynes-Johnson octagon' [13].

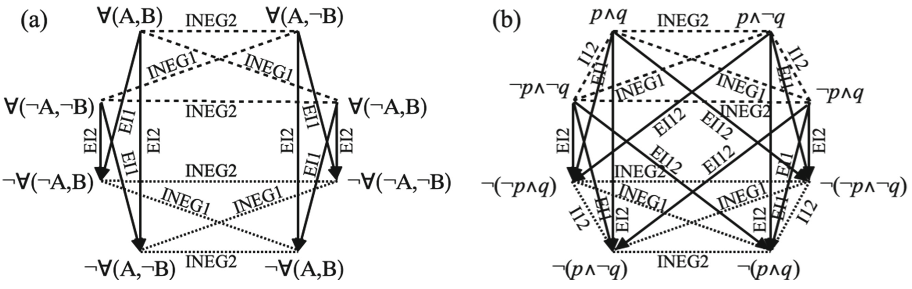


**Fig. 8.** (a) Keynes-Johnson octagon in syllogistics with subject negation—note that $\forall(A, B)$ should be read as *all A are B*; (b) Moretti octagon in propositional logic.

---

[12] In general, for an $n$-ary operator, we have $n + 1$ independent negation positions, viz. 1 external negation and $n$ internal negations (one for each argument position).

Looking at the octagon in Fig. 8(a), we observe, for example, that $\forall(A,B)$ is contrary to two propositions, viz. $\forall(A,\neg B)$ and $\forall(\neg A,B)$. The first of these contrarieties corresponds to an INEG2-relation, and the second one to an INEG1-relation. We also note that the INEG2 of $\forall(A,B)$ is $\forall(A,\neg B)$ and that the INEG2 of $\neg\forall(A,\neg B)$ is $\neg\forall(A,B)$; the first of these INEG2-relations corresponds to a contrariety, and the second to a subcontrariety. The entire distribution of Aristotelian and duality relations in the Keynes-Johnson octagon for syllogistics with subject negation is summarized by the ADM in Fig. 9.

**Fig. 9.** ADM for the Keynes-Johnson octagon in syllogistics with subject negation.

This ADM shows that the correspondence between Aristotelian and duality relations in the Keynes-Johnson octagon is quite regular. Apart from $CD$, all Aristotelian relations correspond to multiple duality relations. Vice versa, however, only INEG1 and INEG2 correspond to multiple Aristotelian relations—all other duality relations correspond to a unique Aristotelian relation. The ADM has 5 connected components, two of which are $\{CD, \textsc{eneg}\}$ and $\{EQ, \textsc{id}\}$, which represent clear-cut correspondences between Aristotelian and duality relations. In comparison with the ADM for the square (cf. Fig. 2), the component $\{SA, \textsc{dual}\}$ from the square has expanded into $\{SA, \textsc{ei}1, \textsc{ei}2\}$. The generalized Post duality relations EI1 and EI2 thus jointly fulfill the role of the original DUAL-relation, in corresponding to $SA$. Similarly, the component $\{C, SC, \textsc{ineg}\}$ has expanded into $\{C, SC, \textsc{ineg}1, \textsc{ineg}2\}$, i.e. the generalized Post duality relations INEG1 and INEG2 thus jointly fulfill the role of the original INEG-relation, in corresponding to $C$ and $SC$.

### 5.2 The Moretti Octagon in Propositional Logic

Another example of generalized Post duality can be observed in the octagon in Fig. 8(b). This octagon is a natural extension of the propositional logic square in Fig. 1(c). It was first studied by Moretti [34] and later also by others [32]. Within the classification of Aristotelian diagrams, it is called a 'Moretti octagon'.

Looking at the octagon in Fig. 8(b), we observe, for example, that $p \wedge q$ is contrary to three propositions, viz. $\neg p \wedge q$, $p \wedge \neg q$ and $\neg p \wedge \neg q$. The first of these contrarieties corresponds to an INEG1-relation, the second one to an INEG2-relation, and the third one to an I12-relation. We also note that the INEG1 of $p \wedge q$ is $\neg p \wedge q$ and that the INEG1 of $\neg(\neg p \wedge q)$ is $\neg(p \wedge q)$; the first of these INEG1-relations corresponds to a contrariety, and the second to a subcontrariety. The entire distribution of Aristotelian and duality relations in the Moretti octagon for propositional logic is summarized by the ADM in Fig. 10.

**Fig. 10.** ADM for the Moretti octagon in propositional logic.

This ADM shows that the correspondence between Aristotelian and duality relations in the Moretti octagon is quite regular. Apart from *CD* and *EQ*, all Aristotelian relations correspond to multiple duality relations. Vice versa, however, only INEG1, INEG2 and I12 correspond to multiple Aristotelian relations—all other generalized Post duality relations correspond to a unique Aristotelian relation. The ADM has 4 connected components, two of which are $\{CD, \text{ENEG}\}$ and $\{EQ, \text{ID}\}$, which represent clear-cut correspondences between Aristotelian and duality relations. In comparison with the ADM for the square (cf. Fig. 2), the component $\{SA, \text{DUAL}\}$ from the square has expanded into $\{SA, \text{EI1}, \text{EI2}, \text{EI12}\}$. The generalized Post duality relations EI1, EI2 and EI12 thus jointly fulfill the role of the original DUAL-relation, in corresponding to *SA*. Similarly, the component $\{C, SC, \text{INEG}\}$ from the square has expanded into $\{C, SC, \text{INEG1}, \text{INEG2}, \text{I12}\}$, i.e. INEG1, INEG2 and I12 thus jointly fulfill the role of the original INEG-relation, in corresponding to *C* and *SC*.

## 6 Conclusion

In this paper we have analyzed the correspondence between Aristotelian and duality relations in four octagons of oppositions. These octagons are all natural extensions of the square of opposition from both an Aristotelian and a duality perspective, and hence, they provide a solid basis for our comparative analysis. The results we obtained are quite nuanced.

On the one hand, the clear-cut correspondence between Aristotelian and duality relations that is found in many squares of opposition (cf. Fig. 2) is lost. In each octagon, we find several cases of a single Aristotelian relation corresponding to multiple duality relations, and vice versa (cf. Figs. 6, 7, 9 and 10). Furthermore, there is no uniform correspondence at the level of *diagrams* either: composed operator duality corresponds to (at least) two types of Aristotelian diagrams (viz. a Buridan octagon and a Lenzen octagon), and generalized Post duality also corresponds to (at least) two types of Aristotelian diagrams (viz. a Keynes-Johnson octagon and a Moretti octagon).

On the other hand, at a higher level of abstraction, the correspondence seems to remain largely intact. Recall that the ADM of the square has 4 connected components. In the ADMs of the Lenzen, Keynes-Johnson, and Moretti octagons, the number of connected components does not decrease. Furthermore, the connected components remain logically meaningful. For example, in the square, *SA* corresponds to DUAL, but in the Lenzen octagon, this Aristotelian relation corresponds to EI, MI and EM, in the Keynes-Johnson octagon to EI1 and EI2, and in the Moretti octagon to EI1, EI2 and EI12. Finally, note that the ADMs for the Lenzen and Moretti octagons (Figs. 7 and 10) are *isomorphic* to each other.

In future work, we will further investigate the correspondence between Aristotelian and duality diagrams. The results obtained in this paper will provide valuable input for such an investigation. Another research question is of a more historical nature. Apart from the square of opposition, the two oldest Aristotelian diagrams that have ever been used, are probably the Buridan octagon (14th century) and the Keynes-Johnson octagon (end of the 19th century). By contrast, the JSB hexagon—which is the most natural extension of the square from a strictly Aristotelian perspective—was only proposed in the 1950s. In this paper, we have argued that, unlike the JSB hexagon, the Buridan octagon and the Keynes-Johnson octagon can also be seen as duality diagrams (according to a suitably generalized notion of duality). Consequently, one might wonder whether these historical facts should primarily be explained in terms of the octagons' duality relations, rather than their Aristotelian relations.

**Acknowledgements.** We would like to thank Margaux Smets and three anonymous reviewers for their valuable feedback. The first author holds a Postdoctoral Fellowship of the Research Foundation–Flanders (FWO).

## References

1. van Benthem, J.: Linguistic universals in logical semantics. In: Zaefferer, D. (ed.) Semantic Universals and Universal Semantics, pp. 17–36. Foris (1991)
2. Blanché, R.: Sur l'opposition des concepts. Theoria **19**, 89–130 (1953)
3. Chow, K.F.: General patterns of opposition squares and 2n-gons. In: Béziau, J.Y., Jacquette, D. (eds.) Around and Beyond the Square, pp. 263–275. Springer, Basel (2012). https://doi.org/10.1007/978-3-0348-0379-3_18
4. Copenhaver, B.P., Normore, C.G., Parsons, T. (eds.): Peter of Spain, Summaries of Logic. Text, Translation, Introduction and Notes. Oxford University Press, Oxford (2014)
5. D'Alfonso, D.: The square of opposition and generalized quantifiers. In: Béziau, J.Y., Payette, G. (eds.) Around and Beyond the Square of Opposition, pp. 219–227. Springer, Basel (2012). https://doi.org/10.1007/978-3-0348-0379-3_15
6. Demey, L.: Algebraic aspects of duality diagrams. In: Cox, P., Plimmer, B., Rodgers, P. (eds.) Diagrams 2012. LNCS (LNAI), vol. 7352, pp. 300–302. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31223-6_32
7. Demey, L.: Structures of oppositions for public announcement logic. In: Béziau, J.Y., Jacquette, D. (eds.) Around and Beyond the Square of Opposition, pp. 313–339. Springer, Basel (2012). https://doi.org/10.1007/978-3-0348-0379-3_22
8. Demey, L.: Interactively illustrating the context-sensitivity of Aristotelian diagrams. In: Christiansen, H., Stojanovic, I., Papadopoulos, G.A. (eds.) CONTEXT 2015. LNCS (LNAI), vol. 9405, pp. 331–345. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25591-0_24
9. Demey, L.: The logical geometry of Russell's theory of definite descriptions. Unpublished manuscript (2017)

10. Demey, L., Smessaert, H.: The relationship between Aristotelian and Hasse diagrams. In: Dwyer, T., Purchase, H., Delaney, A. (eds.) Diagrams 2014. LNCS (LNAI), vol. 8578, pp. 213–227. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44043-8_23
11. Demey, L., Smessaert, H.: Duality in logic and language. In: Fieser, J., Dowden, B. (eds.) Internet Encyclopedia of Philosophy. University of Tennessee, Knoxville (2016)
12. Demey, L., Smessaert, H.: The interaction between logic and geometry in Aristotelian diagrams. In: Jamnik, M., Uesaka, Y., Elzer Schwartz, S. (eds.) Diagrams 2016. LNCS (LNAI), vol. 9781, pp. 67–82. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42333-3_6
13. Demey, L., Smessaert, H.: Metalogical decorations of logical diagrams. Log. Univers. **10**, 233–292 (2016)
14. Demey, L., Smessaert, H.: Shape heuristics in Aristotelian diagrams. In: Kutz, O., Borgo, S., Bhatt, M. (eds.) Shapes 3.0, vol. 1616, pp. 35–45. CEUR-WS (2016)
15. Demey, L., Smessaert, H.: Combinatorial bitstring semantics for arbitrary logical fragments. J. Philos. Log. (2017). https://doi.org/10.1007/s10992-017-9430-5
16. Demey, L., Smessaert, H.: Logical and geometrical distance in polyhedral Aristotelian diagrams in knowledge representation. Symmetry **9**, 204 (2017)
17. Demey, L., Steinkrüger, P.: De logische geometrie van Johannes Buridanus' modale achthoek. Tijdschrift voor Filosofie **79**, 217–238 (2017)
18. Diestel, R.: Graph Theory. Springer, Heidelberg (2006). https://doi.org/10.1007/978-3-662-53622-3
19. Dziewicki, M.H. (ed.): Johannis Wyclif, Tractatus de Logica, vol. 1. Trübner (1893)
20. Hacker, E.A.: The octagon of opposition. Notre Dame J. Form. Log. **16**, 352–353 (1975)
21. Hess, E.: The open future square of opposition: a defense. Sophia **56**, 573–587 (2017)
22. Horn, L.R.: A Natural History of Negation. University of Chicago Press, Chicago (1989)
23. Hughes, G.E., Cresswell, M.J.: A New Introduction to Modal Logic. Routledge, Abingdon (1996)
24. Humberstone, L.: The Connectives. MIT Press, Cambridge (2011)
25. Jacoby, P.: A triangle of opposites for types of propositions in Aristotelian logic. New Scholasticism **24**, 32–56 (1950)
26. Johnson, W.: Logic. Part I. Cambridge University Press, Cambridge (1921)
27. Keynes, J.N.: Studies and Exercises in Formal Logic. MacMillan, London (1884)
28. Klima, G. (ed.): John Buridan, Summulae de Dialectica. Yale University Press, New Haven (2001)
29. Kretzmann, N.: William of Sherwood's Introduction to Logic. Minnesota Archive Editions (1966)
30. Lenzen, W.: How to square knowledge and belief. In: Béziau, J.Y., Jacquette, D. (eds.) Around and Beyond the Square of Opposition, pp. 305–311. Springer, Basel (2012). https://doi.org/10.1007/978-3-0348-0379-3_21
31. Libert, T.: Hypercubes of duality. In: Béziau, J.Y., Jacquette, D. (eds.) Around and Beyond the Square of Opposition, pp. 293–301. Springer, Basel (2012). https://doi.org/10.1007/978-3-0348-0379-3_20
32. Luzeaux, D., Sallantin, J., Dartnell, C.: Logical extensions of Aristotle's square. Log. Univers. **2**, 167–187 (2008)
33. Mikhail, J.: Elements of Moral Cognition. Cambridge University Press, Cambridge (2011)

34. Moretti, A.: The geometry of logical opposition. Ph.D. thesis, Neuchâtel (2009)
35. Moretti, A.: Why the logical hexagon? Log. Univers. **6**, 69–107 (2012)
36. Parsons, T.: The traditional square of opposition. In: Zalta, E.N. (ed.) Stanford Encyclopedia of Philosophy, Summer 2017 edn. CSLI, Stanford (2017)
37. Peters, S., Westerståhl, D.: Quantifiers in Language and Logic. Oxford University Press, Oxford (2006)
38. Read, S.: John Buridan's theory of consequence and his octagons of opposition. In: Béziau, J.Y., Jacquette, D. (eds.) Around and Beyond the Square of Opposition, pp. 93–110. Springer, Basel (2012). https://doi.org/10.1007/978-3-0348-0379-3_6
39. Schumann, A.: On two squares of opposition: the Leśniewski's style formalization of synthetic propositions. Acta Analytica **28**, 71–93 (2013)
40. Sesmat, A.: Logique II. Les Raisonnements. La syllogistique. Hermann (1951)
41. Smessaert, H.: The classical Aristotelian hexagon versus the modern duality hexagon. Log. Univers. **6**, 171–199 (2012)
42. Smessaert, H., Demey, L.: Logical and geometrical complementarities between Aristotelian diagrams. In: Dwyer, T., Purchase, H., Delaney, A. (eds.) Diagrams 2014. LNCS (LNAI), vol. 8578, pp. 246–260. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44043-8_26
43. Smessaert, H., Demey, L.: Logical geometries and information in the square of opposition. J. Logic Lang. Inform. **23**, 527–565 (2014)
44. Smessaert, H., Demey, L.: Visualising the boolean algebra $\mathbb{B}_4$ in 3D. In: Jamnik, M., Uesaka, Y., Elzer Schwartz, S. (eds.) Diagrams 2016. LNCS (LNAI), vol. 9781, pp. 289–292. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42333-3_26
45. Smessaert, H., Demey, L.: Duality patterns in 2-PCD fragments. South Am. J. Log. (2017)
46. Smessaert, H., Demey, L.: The unreasonable effectiveness of bitstrings in logical geometry. In: Béziau, J.Y., Basti, G. (eds.) The Square of Opposition: A Cornerstone of Thought, pp. 197–214. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-45062-9_12
47. Westerståhl, D.: Classical vs. modern squares of opposition, and beyond. In: Béziau, J.Y., Payette, G. (eds.) The Square of Opposition. A General Framework for Cognition, pp. 195–229. Peter Lang (2012)
48. Yao, Y.: Duality in rough set theory based on the square of opposition. Fundamenta Informaticae **127**, 49–64 (2013)

# Syllogistic with Jigsaw Puzzle Diagrams

José Martín Castro-Manzano(✉)

Faculty of Philosophy and Humanities, UPAEP, 21 sur 1103, Puebla, México
`josemartin.castro@upaep.mx`

**Abstract.** In this contribution we introduce a system that represents a modern version of syllogistic by exploiting an analogy with jigsaw puzzles.

**Keywords:** Diagrammatic reasoning · Knowledge representation
Term logic

## 1 Introduction

In this contribution we try to accomplish two goals: *(i)* to introduce a system that represents syllogistic by exploiting an analogy with jigsaw puzzles and *(ii)* show that it is sound and complete with respect to its decision procedure. To reach these goals we briefly review a relation between logic and diagrams (Sect. 2), we mention some general aspects of syllogistic and jigsaw puzzles (Sect. 3), and finally, we present the system along with some of its properties (Sects. 4 and 5).

## 2 Logic and Diagrams

### 2.1 Logical Systems

*Reasoning* is a process that produces information given previous data by following certain norms that allow us to describe *inference* as the unit of measurement of reasoning: inference may be more or less (in)correct depending on the compliance or violation of such norms. *Logical systems*, the tools used to model and better understand inference, may be defined by pairs of the form $\langle L, B \rangle$, where $L$ stands for a language, and $B$ for a semantic base (often equivalent to a calculus). Usually, some syntax is used to determine, uniquely and recursively, the well formed expressions of the system; while semantics is used to provide meaning to such expressions.

### 2.2 Diagrams

In order to represent knowledge, we use internal and external representations. Internal representations convey mental images, for example; while external representations include physical objects on paper, blackboards, or computer

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 657–671, 2018.
https://doi.org/10.1007/978-3-319-91376-6_58

screens. External representations can be further divided into two kinds: sentential and diagrammatic [1]. *Sentential representations*, as the name indicates, are sequences of sentences in a particular language. *Diagrammatic representations*, on the other hand, are sequences of diagrams that contain information stored at one particular *locus*, including information about relations with the adjacent *loci*: diagrams are information graphics that index information by location on a plane [1,2]. For the purposes of this contribution, we take it that a *logic diagram* is a two-dimensional geometric figure with spatial relations that are isomorphic with the structure of logical statements [3, p. 28]. The difference between diagrammatic and sentential representations is that, due to this spatial feature, the former preserve explicitly information about topological relations, while the latter do not—although they may, of course, preserve other kinds of relations. This spatial feature provides some computational advantages: diagrams group together information avoiding large amounts of search, they automatically support a large number of perceptual inferences, and they grant the possibility of applying operational constraints (like *free rides* and *overdetermined alternatives* [4]) to allow the automation of perceptual inference [1].

### 2.3 Logic with Diagrams

To wrap all this up, if reasoning is a process that produces information given previous data and information can be represented diagrammatically, it is not uncomfortable to suggest that diagrammatic inference is the unit of measurement of diagrammatic reasoning: *diagrammatic inference* would be (in)correct depending on the compliance or violation of certain norms. This relation would define our intuitions around the informal notion of *visual inference* and would follow, *ex hypothesi*, classical structural norms (reflexivity, monotonicity, and cut) by way of free rides, that is to say, by way of processes that allow us to gain information without following any step specifically designed to gain it, i.e., processes that allow some reasoner to reach automatically (and sometimes inadvertently) a diagrammatic representation of a conclusion from a given diagrammatic representation of premises [4, p. 32][5]. VENN is a diagrammatic logical system of this sort [6].

## 3 Syllogistic and Jigsaw Puzzles

### 3.1 General Aspects of Syllogistic

*Syllogistic* is a term logic that has its origins in Aristotle's *Prior Analytics* [7] and deals with the consequence relation between categorical propositions. A *categorical proposition* is a proposition composed by two terms, a quantity, and a quality. The subject and the predicate of a proposition are called *terms*: the term-schema S denotes the subject term of the proposition and the term-schema P denotes the predicate. The *quantity* may be either universal (*All*) or particular (*Some*) and the *quality* may be either affirmative (*is*) or negative (*is not*).

These categorical propositions have a *type* denoted by a label (either a (universal affirmative, SaP), e (universal negative, SeP), i (particular affirmative, SiP), or o (particular negative, SoP)) that allows us to determine a *mood.* A *categorical syllogism*, then, is a sequence of three categorical propositions ordered in such a way that two propositions are premises and the last one is a conclusion. Within the premises there is a term that appears in both premises but not in the conclusion. This particular term, usually denoted with the term-schema M, works as a link between the remaining terms and is known as the middle term. According to the position of this last term, four *figures* can be set up in order to encode the valid syllogistic moods or patterns (Table 1)[1].

**Table 1.** Valid syllogistic moods

| Figure 1 | Figure 2 | Figure 3 | Figure 4 |
|---|---|---|---|
| aaa (*Barbara*) | eae (*Cesare*) | iai (*Disamis*) | aee (*Calemes*) |
| eae (*Celarent*) | aee (*Camestres*) | aii (*Datisi*) | iai (*Dimatis*) |
| aii (*Darii*) | eio (*Festino*) | oao (*Bocardo*) | eio (*Fresison*) |
| eio (*Ferio*) | aoo (*Baroco*) | eio (*Ferison*) | |

## 3.2 General Aspects of Jigsaw Puzzles

A *tessellation* in the euclidean plane is a covering of the plane without gaps or overlappings by congruent polygons called tiles. A square-tiling, for example, is a tessellation that uses squares as tiles; formally, it is a subset of the euclidean plane that is the union of two sets of equally spaced parallel lines such that the lines of each different set are perpendicular [8, p. 69] (Fig. 1a). Notable and more complex examples of tessellations can be found within Islamic patterns [9], Kepler's monsters [10], Escher's lithographs [11], and Penrose's aperiodic tessellations [12].

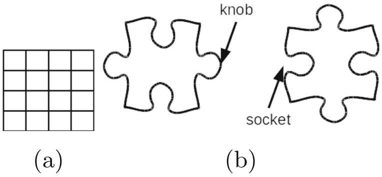


**Fig. 1.** Tessellations

---

[1] For sake of brevity, but without loss of generality, here we omit the syllogisms that require existential import.

In particular, a *jigsaw puzzle* is some sort of tessellation: it is a tiling array composed by a finite set of tessellating pieces that require assembly by way of the interlocking of tiles known as *knobs* and *sockets* (Fig. 1b). These types of puzzles date back as far as Archimedes [13, p. 13], although the typical pictorial jigsaw puzzles we are familiar with have their roots in the 1760's when John Spilsbury first dissected maps for educational purposes [14, p. 24].

## 4 The System JGSW

With this background in mind, we now suggest that a modern version of syllogistic can be represented with jigsaw puzzles: just as jigsaw puzzles require the *interlocking of tiles*, syllogisms require the *linking of terms*. JGSW is a diagrammatic system that exploits this analogy by using a square-tiling tessellation in order to provide representation and a decision procedure for syllogistic. In this section we define JGSW by detailing its vocabulary, its syntax, and semantics.

### 4.1 Elements of JGSW

The vocabulary of JGSW is defined by two elementary diagrams (i.e., pieces or tiles), sockets and knobs (Fig. 2).

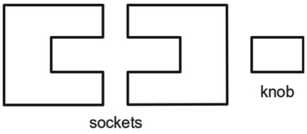
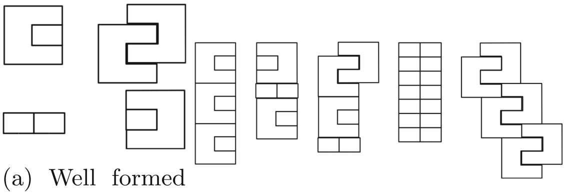


**Fig. 2.** Vocabulary

Syntax is defined by two rules: *(i)* given two elementary diagrams, the combinations of diagrams in Fig. 3a are well formed diagrams (wfd); and *(ii)* a finite sequence of wfds is also a wfd (we call this sequence a *stack*) (Fig. 3b).



(a) Well formed diagrams

(b) Stacks

**Fig. 3.** Syntax

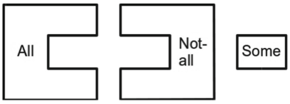
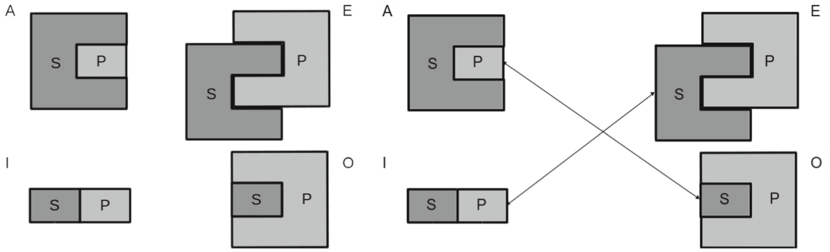


**Fig. 4.** Semantics

Semantics is given by the interpretation of Fig. 4.

With these elements we can represent the categorical propositions. For sake of brevity, we will label each tile with an affirmative subject or predicate term-schema, S or P, given that the tile already indicates the quantity associated to each term (later we will use $\overline{S}$ and $\overline{P}$ in order to denote the complementary term-schemas of S or P); and for sake of visualization, we will color the terms[2]. The diagrams in Fig. 5a represent, thus, the categorical propositions: (A) SaP, (E) SeP, (I) SiP, and (O) SoP.

With this information, we can arrange a boolean *square of opposition* in which the only rules preserved would be the rules for contradiction between A and O, and between E and I; contrariety, subalternation, and subcontrariety do not work (Fig. 5b). Thus, this system's square behaves as a modern square of opposition rather than a traditional square. However, despite this apparent shortcoming, the equivalence rules of conversion, contraposition, and obversion are preserved in JGSW by the mechanical operations of rotating diagrams and switching tiles.



(a) Categorical propositions (b) Square of opposition

**Fig. 5.** Propositions

### 4.2 Equivalence Rules

*Conversion* consists in rotating a wfd by 180°. In Fig. 6 we can notice the conversion of a proposition E (I) semantically—and visually—remains a proposition E (I), but the conversion of a proposition A (O) does not remain a proposition A (O). Hence, conversion holds only between propositions E and I (the arrows in Fig. 6 indicate the equivalence).

---

[2] We use colors with the purpose of showing the reasoning process. A logical use of color can be seen in [15].

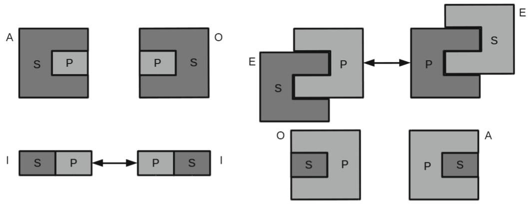


**Fig. 6.** Conversion

*Contraposition* consists in switching the position of a socket to a knob (or vice versa) with its respective complementary terms. In Fig. 7 we can see the application of a contraposition to a proposition A (O) remains semantically and visually a proposition A (O), but contraposition of propositions E and I produce propositions I and E, which are not equivalent to the original propositions. Thus, contraposition holds only between propositions A and O.

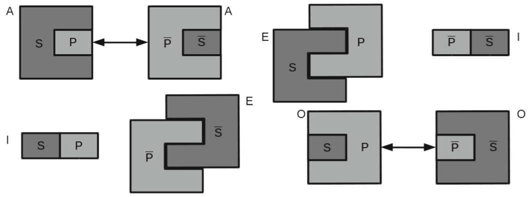


**Fig. 7.** Contraposition

Finally, *obversion* consists in switching sockets to knobs (or vice versa) just in the predicate terms along with their respective complements. In Fig. 8 we can see the application of an obversion to any proposition produces an equivalent proposition.

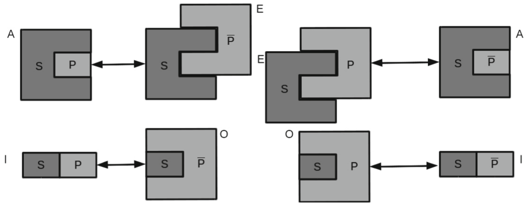


**Fig. 8.** Obversion

### 4.3 Decision Procedure

Up to this point, we can represent categorical propositions. Now we need a way to represent categorical syllogisms and some method to decide whether these are (in)valid. In order to achieve the former goal, we just need to stack up the wfds that represent the premises of each syllogism; to achieve the latter, suppose we build categorical propositions using a single term-schema, say M (Fig. 9). We can observe that, from these representations, only proposition A, *All* M *is* M, is a tautology. Using this tautology, we suggest a decision procedure for JGSW that takes any syllogism $\sigma$ as an input and decides whether it is (in)valid by verifying a single rule: if the interlocking of its middle terms produces a proposition of type A, the syllogism is valid; otherwise, it is invalid (Algorithm 1).

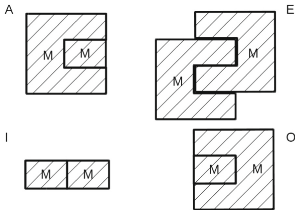


**Fig. 9.** Propositions using the term M

**Algorithm 1.** $\mathcal{A}^1$

```
Input: syllogism $\sigma$
if interlocking of middle terms of $\sigma$ == A then
    $\sigma$ is valid;
else
    $\sigma$ is invalid;
end
```

### 4.4 Validity of the Syllogisms

Using the rules of equivalence and the previous decision procedure, we now prove the 15 valid syllogisms depicted in Table 1 (Figs. 10, 11, 12 and 13). We start by stacking up the wfds that represent the premises of each syllogism. Then we apply $\mathcal{A}^1$ and we check if the middle term tiles interlock each other forming a proposition A (a step denoted by the arrows in the following Figures); in case it does, the inference is valid, thus allowing a free ride by letting the tiles S and P interlock in the third diagram (i.e., the conclusion), which is below the turnstile. These steps will be exemplified with more detail below; meanwhile, in what follows, conv stands for conversion, contra for contraposition, and obv for obversion.

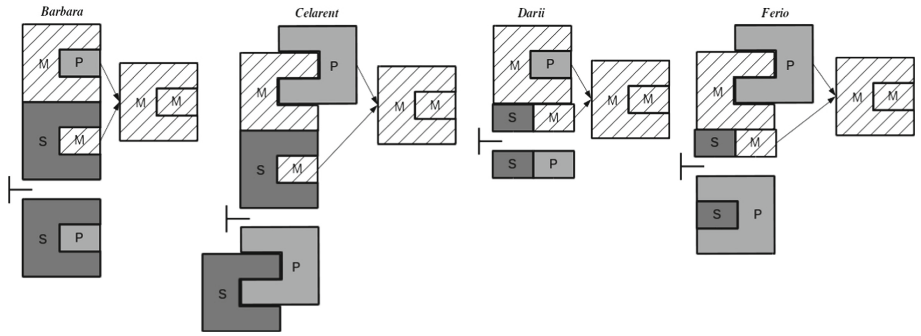


**Fig. 10.** Proofs of the syllogisms of the first figure

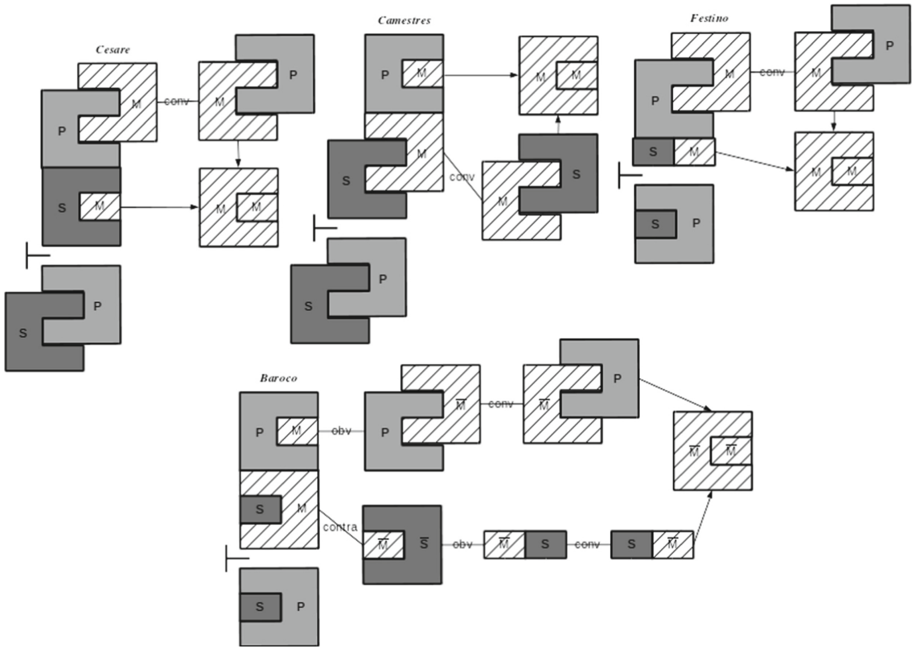


**Fig. 11.** Proofs of the syllogisms of the second figure

### 4.5 Examples

Now, in order to illustrate how we can use JGSW, let us show some examples that give the feeling of the mechanical operations involved, because besides having diagrammatic properties, JGSW can be materialized, so to speak: JGSW is, after all, a jigsaw puzzle made up of physical tiles.

Hence, the following examples differ from the previous proofs in that these are not proofs but applications that presuppose the results of the proofs. First we show an example of a valid syllogism, then the result of applying our method

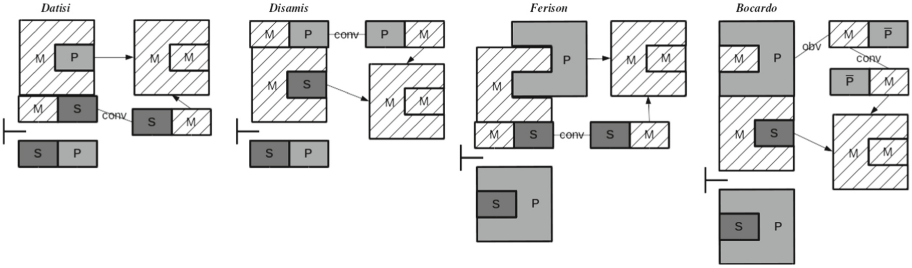


**Fig. 12.** Proofs of the syllogisms of the third figure

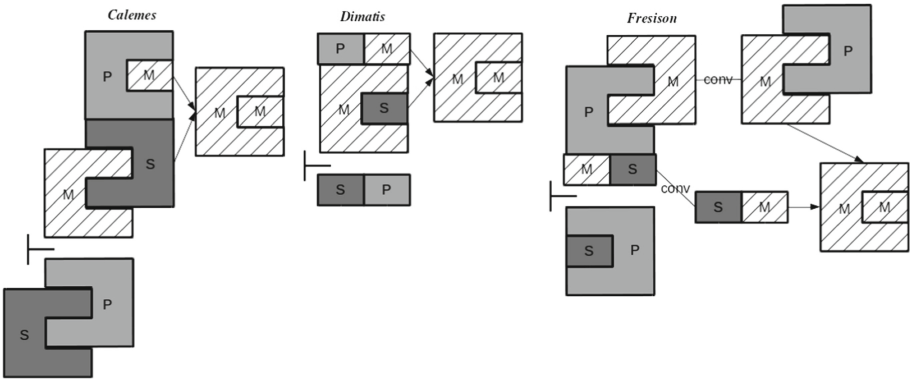


**Fig. 13.** Proofs of the syllogisms of the fourth figure

to an invalid syllogistic form, and finally, an example of how the system may be used to find a missing conclusion given an enthymeme. In the next Figures, the arrows indicate transitions between steps and the superscript index denote steps.

**Example 1.** *Consider a Baroco syllogism (Fig. 14). We stack up the premises (step 1) and we overlap them in order to meet the* S *with the* P*, because the conclusion requires meeting the* S *with the* P *(step 2). Since the* P *must be to the right in the conclusion, we rotate the overlapped diagrams (step 3) and we can observe that, by an application of* $\mathcal{A}^1$ *, since the middle terms do produce a proposition* A *(step 4), this example must be a valid syllogism, and thus, we can reach the conclusion (step 5). This is a fair example of a free ride.*

Notice how the proof of the *Baroco* syllogism (Cf. Fig. 11) differs from the application of JGSW to it: while the proof requires the adherence to the explanatory principles of syllogistic (semantics, equivalence rules, and so on), the application does not appeal to such principles; therefore, we do not require a previous knowledge of what the conclusion should be. Also, note that the operation of *overlapping* resolves a potential issue with $\mathcal{A}^1$: it is true that the algorithm merely

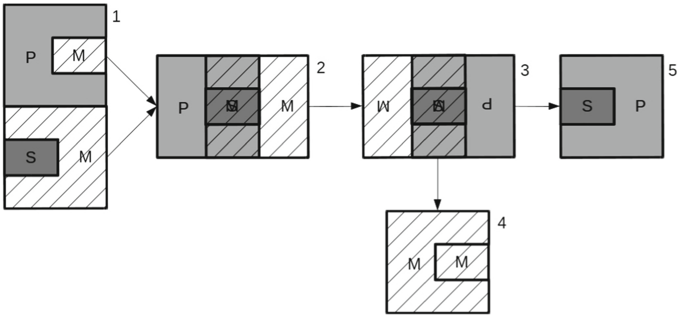


**Fig. 14.** An example of a valid syllogism

checks whether the middle terms produce a proposition A when the premises are stacked, which seems to imply that only the syllogisms from Fig. 1 will come out to be valid; and if the equivalence rules must also be applied, and it is only after they are applied that $\mathcal{A}^1$ will give the correct decision, it seems there is no clear indication that the application of rules will ever terminate. However, the process of overlapping, as we have exemplified it in the previous example, allows us to dodge this issue because it results in a configuration that grants a direct application of $\mathcal{A}^1$. This solution yields Algorithm 2.

**Algorithm 2.** $\mathcal{A}^2$

```
Input: syllogism $\sigma$
$\sigma' \leftarrow$ overlap $\sigma$
apply $\mathcal{A}^1(\sigma')$
if $\mathcal{A}^1(\sigma') == invalid$ then
    $\sigma'' \leftarrow$ rotate $\sigma'$;
    apply $\mathcal{A}^1(\sigma'')$;
end
```

**Example 2.** *Consider the syllogistic form* iai-*1 (Fig. 15). We first stack up the premises (step 1) and we overlap them in order to meet the* S *with the* P *(step 2). By* $\mathcal{A}^2$*, since the middle terms do not produce a proposition of type* A *(step 3), this example must be invalid, that is to say, we cannot reach the conclusion: this is indicated by the* X.

Notice that, in this particular example, regardless of how the original stack is manipulated, we will never reach the criterion of validity, since it can be shown that a proposition of type I (i.e. two connected knobs) cannot be transformed into a proposition of type A (i.e. a socket linked to a knob). A general answer to this issue will be given with the results of Sect. 5.

**Example 3.** *Consider a syllogistic form with missing conclusion:* ia?-*3 (Fig. 16). Since the application of* JGSW *differs from a proof, we do not require*

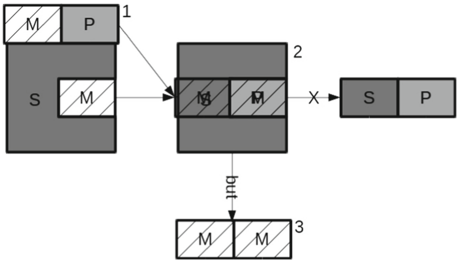


**Fig. 15.** An example of an invalid syllogism

*to know the right conclusion before we begin. Thus, if the premises yield a conclusion, the application of* JGSW *must provide such conclusion. So, we stack up the premises (step 1) and we apply $\mathcal{A}^2$. Since the middle terms do produce a proposition of type* A *(step 3), we can find the right and unique conclusion (step 4). This process indicates that this example has the form* iai*-3: it is a Disamis syllogism. The same method can be applied when searching for missing premises.*

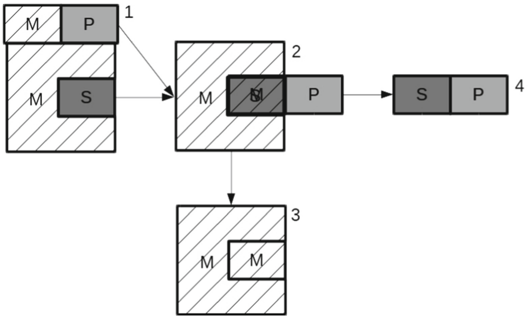


**Fig. 16.** An example of an enthymeme

## 5 Properties of JGSW

In this section we try to accomplish our second goal by exploring the soundness and completeness of JGSW w.r.t. $\mathcal{A}^2$. But before we show proof of these properties, we need a preliminary result that we call Aristotle's lemma, in homage to *Pr. An.* A.1, 25b1.

**Lemma 1 (Aristotle's lemma).** *All valid syllogisms can be reduced to valid syllogisms of the first figure.*

*Proof.* To provide proof of this statement, we proceed by cases (Figs. 17, 18 and 19). The proof requires us to introduce two simple rules: move, which moves a

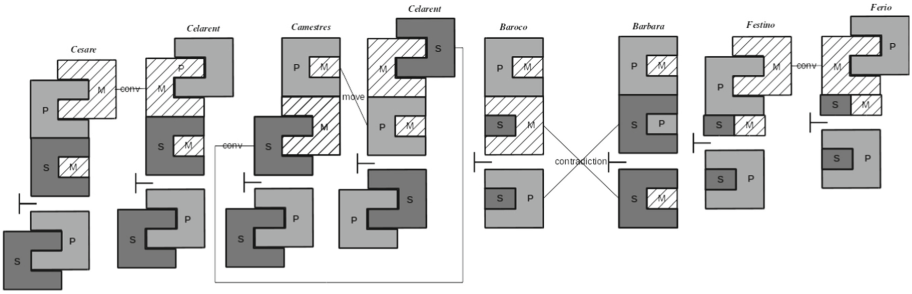


**Fig. 17.** Reduction of the second figure to the first figure

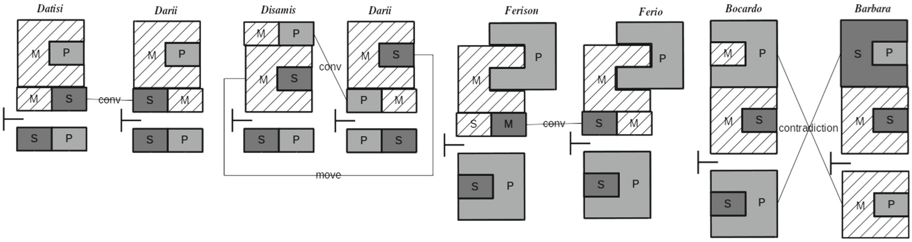


**Fig. 18.** Reduction of the third figure to the first figure

diagram in the stack of premises; and `contradiction`, which takes the contradictory diagram of the conclusion and interchanges it with a diagram in the stack of premises.

With this result in our hands, we can now present proofs of soundness and completeness w.r.t. $\mathcal{A}^2$. Let us denote the application of $\mathcal{A}^2$ to a given syllogism $\sigma_{i,j}$ from figure $i \in \{1, 2, 3, 4\}$ and row $j \in \{1, 2, 3, 4\}$, as in Table 1, by $\mathcal{A}^2(\sigma_{i,j})$; thus, for instance, the application of $\mathcal{A}^2$ to a *Ferison* syllogism is $\mathcal{A}^2(\sigma_{3,4})$; and for sake of exposition, $\mathcal{A}^2(\sigma_{4,4})$ is a placeholder.

**Proposition 1 (Soundness w.r.t. $\mathcal{A}^2$).** *If $\mathcal{A}^2(\sigma) = valid$, then $\sigma$ is valid.*

*Proof.* We prove this proposition by cases. Since there are four figures, we need to cover each valid syllogism from each figure. This is what we have done in Sect. 4, Figs. 10, 11, 12 and 13: for every $\sigma_{i,j}$ we have checked that when $\mathcal{A}^2(\sigma_{i,j}) = valid$, $\sigma_{i,j}$ is valid.

**Proposition 2 (Completeness w.r.t. $\mathcal{A}^2$).** *If $\sigma$ is valid, then $\mathcal{A}^2(\sigma) = valid$.*

*Proof.* We prove this by contradiction. Suppose that for all $i$, $j$, syllogism $\sigma_{i,j}$ is valid but for some valid syllogism $\sigma_{k,j}$, $\mathcal{A}^2(\sigma_{k,j}) = invalid$. We know $\sigma_{i,j}$ is valid and if we apply $\mathcal{A}^2(\sigma_{k,j})$ we obtain $\mathcal{A}^2(\sigma_{k,j}) = valid$, as we have seen from Proposition 1. Since all valid syllogisms $\sigma_{n>1,j}$ can be reduced to the valid syllogisms of Fig. 1 by Aristotle's lemma, it follows that $\mathcal{A}^2(\sigma_{n>1,j}) = valid$,

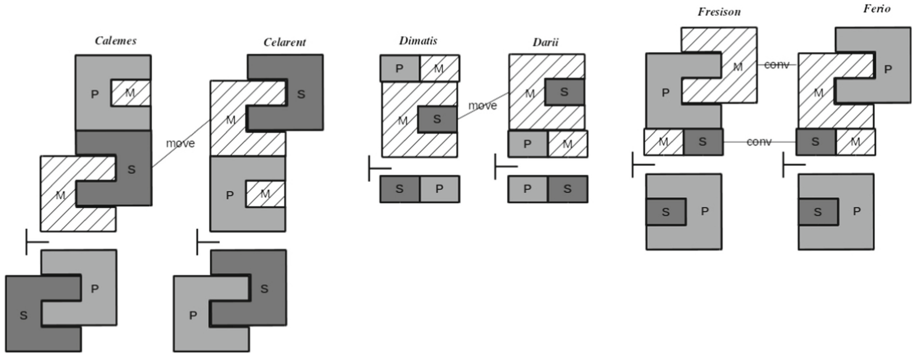


**Fig. 19.** Reduction of the fourth figure to the first figure

and thus, for all valid syllogisms $k$, $\mathcal{A}^2(\sigma_{k,j}) = valid$, which contradicts our assumption.

**Proposition 3 (Decidability).** JGSW *is decidable.*

*Proof.* JGSW is decidable w.r.t. $\mathcal{A}^2$ since it is a finite, sound, and complete procedure that takes any syllogism and decides whether it is (in)valid.

**Proposition 4 (Equivalence).** JGSW *is equivalent to* VENN *w.r.t. a syllogistic base.*

*Proof.* In order to provide proof for this statement we show that every valid syllogism in JGSW is a valid syllogism in VENN and vice versa. From left to right: suppose that for any valid syllogisms $\sigma$, $\sigma_{i,j}$ is a valid syllogism in JGSW but is invalid in VENN. Given the soundness and completeness of JGSW, if $\sigma_{i,j}$ is a valid syllogism in JGSW, $\sigma_{i,j}$ is valid *simpliciter*. But since VENN is sound and complete as well [6], if $\sigma_{i,j}$ is invalid in VENN, then $\sigma_{i,j}$ must be invalid, which is a contradiction. From right to left the proof is similar.

What this brief exploration shows is that JGSW is a *bona fide* diagram-based logic because it produces valid (*soundness*) and only valid inferences (*completeness*) by providing a mechanical method of decision (*decidability*) that helps the automation of perceptual inference while preserving our common intuitions about diagrammatic inference (*equivalence*).

### 5.1 Representational Properties

In order to evaluate the representational attributes of JGSW, let us consider a framework based upon [2, p. 305]. According to it, these representational qualities include: comprehension (diagrams promote system understanding), clarity (diagrams are not ambiguous), parsimony (diagrams are explanatory), relevance (diagrams support problem solving), and separability (diagrams allow multilayered descriptions).

1. *Comprehension.* Does JGSW promotes system understanding? It seems adequate to say that JGSW does promote system understanding, since it offers a structural explanation of categorical propositions (JGSW diagrams are structures that represent terms) and a working explanation of the main core of syllogistic (the algorithms provide an explanation of syllogistic by exploiting an analogy with jigsaw puzzles).
2. *Clarity.* Are JGSW diagrams ambiguous? JGSW diagrams have a clear motivation and a well defined vocabulary. Syntax and semantics guarantee that JGSW diagrams are not syntactically nor semantically overloaded because each diagrammatic object means a thing and one thing only.
3. *Parsimony.* Are JGSW diagrams explanatory? It seems fair to say that JGSW diagrams are parsimonious in so far as they provide a level of abstraction high enough to explain a term logic in which the basic elements include, properly speaking, terms, quantity, and quality.
4. *Relevance.* Do JGSW diagrams support problem solving? It also seems fair to say that JGSW diagrams are good enough to solve inferential problems, given that they allow perceptual inference while preserving our common intuitions about diagrammatic inference.
5. *Separability.* Do JGSW diagrams allow multilayered descriptions? It is clear JGSW diagrams can be (dis)assembled as to avoid representational issues. Plus, the mechanical operations of rotation, interchange, stacking, and interlocking allow modular descriptions of syllogistic inference.

## 6 Conclusions

By exploiting an analogy with jigsaw puzzles, in this contribution we have introduced a diagrammatic logical system that represents a modern version of syllogistic. We think this goal is compelling in and of itself, but it is also interesting because a system like this has not been developed previously, as far as we are aware (Cf. [3,16]); and because the system offers a non-linear (Cf. [17–19]) and non-regional (Cf. [6,20,21]) diagrammatic approach for syllogistic that is capable of being externalized—given the very analogy with jigsaw puzzles—, for instance, by way of wooden or metal tiles, thus allowing combinations of colors, materials, and textures that may provide some sort of non-linguistic information for performing heterogeneous inference. Also, we are currently trying to incorporate another aspects of syllogistic inference with singular terms, relations, and non-classical quantifiers.

## References

1. Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. Cogn. Sci. **11**(1), 65–100 (1987)
2. Nakatsu, R.: Diagrammatic Reasoning in AI. Wiley, Hoboken (2009)
3. Gardner, M.: Logic Machines and Diagrams. University of Chicago Press, Chicago (1982)

4. Shimojima, A.: Operational constraints in diagrammatic reasoning. In: Logical Reasoning with Diagrams. Oxford University Press, Oxford (1996)
5. Shimojima, A.: Semantic Properties of Diagrams and Their Cognitive Potentials. Center for the Study of Language and Information-Lecture Notes Series. CSLI Publications/Center for the Study of Language & Information (2015)
6. Shin, S.: The Logical Status of Diagrams. Cambridge University Press, Cambridge (1994)
7. Aristotle, S.R.: Prior Analytics. Hackett Classics Series. Hackett, Indianapolis (1989)
8. Meurant, G.: Noneuclidean Tessellations and Their Groups. Elsevier Science, New York (1974)
9. Critchlow, K.: Islamic Patterns: An Analytical and Cosmological Approach. Schocken Books, New York City (1976)
10. Kepler, J., Aiton, E., Duncan, A., Field, J.: The Harmony of the World. American Philosophical Society, Philadelphia (1997)
11. Escher, M., Brigham, J.: The Graphic Work of M.C. Escher. Pan Books, London (1978)
12. Penrose, R.: The role of aesthetics in pure and applied mathematical research. Bull. Inst. Math. Appl. **10**, 266–271 (1974)
13. Slocum, J., Botermans, J.: Puzzles Old and New. Seattle University of Washington Press, Seattle (1986)
14. Williams, A., Shortz, W.: The Jigsaw Puzzle: Piecing Together a History. Berkeley Books, New York City (2004)
15. Rini, A.: Aristotle's Modal Proofs: Prior Analytics A8–22 in Predicate Logic. The New Synthese Historical Library. Springer, Netherlands (2010). https://doi.org/10.1007/978-94-007-0050-5
16. Moktefi, A., Shin, S.J.: A history of logic diagrams. In: Gabbay, D.M., Francis Jeffry Pelletier, J.W. (ed.) Handbook of the History of Logic, vol. 11, pp. 611–682. North-Holland (2012)
17. von Leibniz, G., Couturat, L.: Opuscules et fragments inédits de Leibniz: extraits des manuscrits de la Bibliothèque royale de Hanovre. Olms paperback, Olms (1961)
18. Englebretsen, G.: Linear diagrams for syllogisms (with relationals). Notre Dame J. Formal Logic **33**(1), 37–69 (1991)
19. Pagnan, R.: A diagrammatic calculus of syllogisms. J. Logic, Lang. Inf. **21**(3), 347–364 (2012)
20. Venn, J.: On the diagrammatic and mechanical representation of propositions and reasonings. Edinb. Dublin Philos. Mag. J. Sci. (1881). R. Taylor, London
21. Savio, M.: AE (Aristotle-Euler) diagrams: an alternative complete method for the categorical syllogism. Notre Dame J. Formal Logic **39**(4), 581–599 (1998)

# Diagrammatically Formalising Constraints of a Privacy Ontology

Shan Chen and Aidan Delaney(✉)

The University of the South Pacific, Suva, Fiji
shan.chen@usp.ac.fj, aidan@ontologyengineering.org

**Abstract.** Data privacy is a cross-cutting concern for many software projects. We reify a philosophically inspired model for data privacy into a concept diagram. From the concept diagram we extract the privacy constraints and demonstrate one mechanism for translating the constraints into executable software.

## 1 Introduction

There have been many definitions and analyses of "privacy" as a concept; however, clarity and consensus are still lacking. A widely cited definition of privacy is "the right to be left alone" [1]. In the contemporary digital information era, the information aspect of privacy dominates one's privacy: "you are your information" [2]. It follows that, privacy is largely information privacy. The right to privacy (to be left alone) is therefore, the right to information privacy.

The difficulty in maintaining information privacy can be illustrated using an example of transitive friendship on social networks. In the example a student has not shared their phone number with an employer. However, another employee of the same organisation, who is a personal friend of our student, may share our student's phone number with the employer. In this case, the employee has moved the visibility of the student's number from the personal domain to a professional domain. The example serves to further demonstrate how information privacy is not solely a problem of software security. To address the need for a model for privacy the 3CR model was developed [3].

In this paper we present a reification of an extension of 3CR using concept diagrams [4]. As this is the first application of concept diagrams to enforce runtime constraints we also believe the method we followed is of use to researchers. Finally, we demonstrate how key constraints from the concept diagram can be implemented in a software application.

This paper is organised as follows. The 3CR model is presented in Sect. 2. A brief introduction to concept diagrams is presented in Sect. 3. Our method for reifying 3CR is presented in Sect. 4 and our constraint implementation, using Haskell's software transactional memory, is explained. We present a discussion on the related work in Sect. 5. Finally, the conclusion and future work are presented in Sect. 6.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 672–679, 2018.
https://doi.org/10.1007/978-3-319-91376-6_59

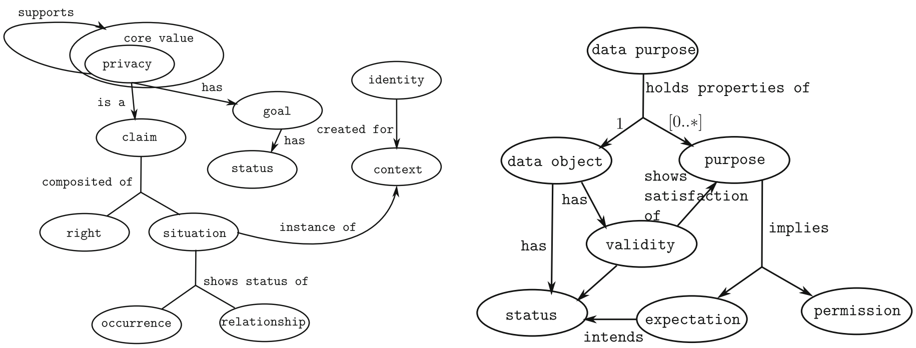
## 2 3CR: A Privacy Ontology

The 3CR model is grounded in the core value framework [5] and is presented in an informal diagrammatic notation, Fig. 1a, which is visually similar to that of concept diagrams. The model concerns itself with the question of an actor's agency regarding their information or "*who* can *do what* to my information".

The *who* concerns one's relationship to others. This concern requires a consideration of situations in which the relationship is recognized. Since relationships evolve over time, a situation is determined spatiotemporally - i.e., when and where, if applicable. Thus, the right to information privacy has a situation dimension - i.e., in what situation can the right be claimed.

The *what* concerns information content and details its information about "my information". The right in this dimension is to determine size, volume and granularity of the allowed operations on the information.

The *do* concerns actions on the *what* (i.e., the selected information). Do-actions can be observation, presentation, access, manipulation or distribution. Observation means watching and remarking on the information, presentation refers to the freedom in presenting the information (i.e., when, where and how to present the information), access means ways of viewing and retrieving the information, manipulation refers to modification of the information, and distribution means sharing the information out with one or among a number of recipients.



(a) (b)

**Fig. 1.** Ontological grounding of 3CR

The 3CR model views the consistency between how data is used and the purposes for which it is collected as fundamental to protecting privacy. This view follows from two principles of information privacy introduced by the OECD[1], namely the *Purpose Specification Principle* and the *Use Limitation Principle*. A UML-like notation is used in [3] to present the *grounding* of 3CR in the purpose

---

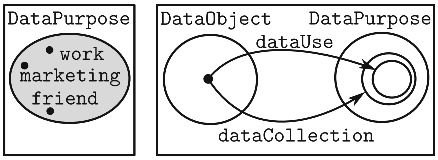
[1] http://www.oecd.org/sti/ieconomy/oecdprivacyframework.pdf.

specification principle, Fig. 1b. The Purpose Specification Principle governs the reasons the data is gathered and restricts further uses of the data to the purposes for which it was collected. We refer to the 3CR model with an informal grounding of data purpose and data usage as the 3CR+ model [3]. The next section introduces concept diagrams that will be used to reify 3CR+ in Sect. 4.

## 3 Concept Diagrams

Concept diagrams [6] are a formal diagrammatic system which focus on expressing semantic web concepts. Concept diagrams represent classes and properties using closed contours and arrows respectively. Concept diagrams express subsumption and disjointedness of classes spatially, using the same visual notation that Euler diagrams use to express subset and disjoint sets. Arrows in a concept diagram represent properties and are allowed between contours and between individuals and contours. The notation for individuals within classes is borrowed from Spider diagrams [7], which denote an individual using a dot as depicted in Fig. 2b.

The example concept-diagram in Fig. 2b asserts the existence of two classes. These are `DataObject` and `DataPurpose`. Moreover, it asserts that elements of `DataObject` are related to subsets of `DataPurpose` under `dataCollection`. In addition, the property `dataUse` relates the same anonymous individual of `DataObject` to some anonymous subset of the "image" of `dataCollection`. That is to say, `dataUse` is a sub-property of `dataCollection`.



(a) (b)

**Fig. 2.** Constraints on data purpose

The semantics of concept diagrams are provided in [6]. Some facts from the diagram in Fig. 2b are captured by the following statement in description logic:

$$DataPurpose \subseteq DataObject \subseteq \bot \sqcap (dataCollection \sqsubseteq dataUse)$$
$$\sqcap DataPurpose : work \sqcap DataPurpose : marketing \sqcap DataPurpose : friend$$

This statement in description logic also provides our motivation for preferring to diagrammatically formalise 3CR+. The stakeholders in the formalisation process were uncomfortable working in symbolic logic but embraced the diagrammatic presentation.

## 4 Reifying and Implementing 3CR+

3CR+ is presented as a set of diagrams; an ontology in a concept-diagram like notation and data use/purpose presented in a UML-like notation. Our method for formalising 3CR+ produces a concept diagram and follows 3 steps:

*Formalising Individuals.* The informal 3CR+ diagrams do not distinguish between individuals and classes. Using concept diagrams we can represent this distinction and introduce more of the information described in the text accompanying 3CR+ into the diagram.

The diagram in Fig. 1a represents privacy using a closed contour. Under a concept diagram interpretation this asserts that privacy is a class. However the text in [3] enumerates the core values as a superset of those contained in the UN declaration of human rights[2]. As such, there are only a limited number of `CoreValues`. To capture both the diagrammatic and textual information we represent each textually enumerated item as a spider in Fig. 2a. The upper bound on the cardinality of the enumeration is denoted using shading.

*Formalising Properties.* We consider the textual description of 3CR+ and the relationships depicted in the existing diagrams. These relationships are formalised as properties with labels that convey their intended semantics more specifically. Moreover, we replace all the ternary relations in 3CR+ with multiple binary relations.

Of particular interest is the interpretation of `data usage` and `data purpose` in Fig. 1b. In both cases there exists a ternary relation involving `data object` and either `purpose` or `usage`. The accompanying text in [8] informs that "usage ... can be interpreted by 'purpose of use' ". As such, this constraint may be modelled in a concept diagram using property subsumption. We introduce the property `dataUse` relating `DataObject` to `DataPurpose`. The `dataUse` property is subsumed by `dataCollection` and reflects the key constraint that data objects must not be used for purposes other than those which they were collected.

*Enriching Properties.* By enriching properties we link the formalised concept-diagram to other ontologies realising some of the promised advantages of the semantic web. The text accompanying Fig. 1a describes how the diagram represents a user's perspective of 3CR+. In a concept diagram we wish to represent the relationship between all users and the `CoreValue` of `privacy`. Using the extensibility of semantic web technologies we may introduce a `Person` class from the FOAF ontology[3] to represent our users . A `Person` may be related to an identity under the `foaf:nick` property. The change in perspective, brought about by the reification process, requires us to realign the properties from `Person` through `Identity` to `CoreValues`. This relates all people to `privacy`.

---

[2] http://www.un.org/en/universal-declaration-human-rights/.

[3] http://xmlns.com/foaf/spec/.

**Constraint Implementation.** To enforce constraints from the formalised 3CR+ ontology we use software transactional memory (STM) techniques from [9]. Our example demonstrates aspects of constraint assertion that cannot be modelled in STM alone and requires the programmer to show due care and attention to protecting privacy. The concept diagrams containing our constraints serve to guide the programmer implementing the system.

We concentrate on the contacts directory in Sekope's mobile phone. Sekope is the friend of our student, from the introdution, and both of them are employed by the same company. We consider two constraints, in Fig. 2, extracted from our formalisation process. In this instance we find that there are at least three `DataPurposes`. We can infer from the context that Sekope stores our student's number for the purpose of `friend` indicating that it is not for use in a `work` purpose. We implement the concept of a `DataPurpose` in Haskell using an algebraic data type on line 1 of Fig. 3. A contact record is also represented as an algebraic data type but in this case there are two functions `dataCollection` and `dataUse` that relate a `ContactEntry` to a set of `DataPurposes`. Haskell's type system, like the type systems of all mainstream programming languages, cannot express the constraint that for every `DataPurpose` the range of `dataUse` must be a subset of `dataCollection`. The constraint expressed in the diagram in Fig. 2b must necessarily be implemented using run-time, as opposed to compile-time, checks.

```
1  data DataPurpose = Friend | Work | Marketing
2  data ContactEntry = ContactEntry {
3    number :: PhoneNumber
4    , dataCollection :: (S.Set DataPurpose), dataUse :: (S.Set DataPurpose)
5  }
```

**Fig. 3.** Implementing `DataPurpose`, `dataCollection` and `dataUse`

We implement a directory of contacts as a transactional variable. Because of this design decision we may use features of Haskell's software transactional memory implementation to specify run-time constraints. The constraint that the use purposes agree with the collection purposes of each `ContactEntry` is expressed on line 5 of Fig. 4. This invariant constraint is installed into the STM subsystem at runtime. The constraint is checked after each transaction involving the transactional variable. If the constraint fails, then the transaction is rolled-back. Therefore inserting 'ContactEntry 1234 (S.fromList [Friend, Work]) (S.fromList [Friend])' will succeed as the set of use purposes (S.fromList [Friend]) is a subset of the collection purposes (S.fromList [Friend, Work]). A transaction to insert an invalid entry 'ContactEntry 5555 (S.fromList [Friend, Work]) (S.fromList [Marketing])' will fail as the set of use purposes is not a subset of the collection purposes. In short, if Sekope collects our student's number for the purposes of being a `friend`, he cannot then add her number to his directory in a `work` context.

The object of our exercise is to ensure that Sekope cannot share his friend's phone number with their employer. This cannot be modelled inside the Haskell

```
predicate_dataUseIsCollectionPurpose :: ContactsDirectory → STM Bool
predicate_dataUseIsCollectionPurpose addressBook = do
  addresses ← getAllAddresses addressBook
  return $ all checkPurposes addresses
  where checkPurposes number =
    S.isSubsetOf (dataUse number) (dataCollection number)
```

**Fig. 4.** Describing an invariant

STM constraint system as it is ill-advised make reading a transactional variable conditional upon the state of the variable itself. We use information hiding to provide a `getContact` function, see Fig. 5, which returns the contact if and only if the stated use purpose for the contact is one of the collection purposes of the contact. Internally in our module we make use of `getContactUnsafe`, a function that reads a `ContactEntry` from our transactional `ContactsDirectory` without checking whether the use purpose agrees with the collection purpose. Haskell's information hiding ensures that we do not expose our *unsafe* function to the rest of the application. As such, the transfer of a friend's phone number is only allowed if the number does not leave the `friend` context.

```
getContactUnsafe :: PhoneNumber → ContactsDirectory →
STM (Maybe ContactEntry)
getContact :: PhoneNumber → (S.Set DataPurpose) →
STM (Either (Maybe ContactEntry))
transferContact :: PhoneNumber → PhoneNumber → ContactsDirectory →
STM (Either (Maybe ContactEntry))
```

**Fig. 5.** Signatures for safely transferring a contact

We have presented our formal reification of the 3CR+ ontology and we have demonstrated how the constraints may be realised in a software application. We now turn our attention to the related work in this field.

## 5 Related Work

KAoS [10] provides a mechanism for specifying ontology-based access policies in a variety of systems including web services. KAoS policies can be specified in OWL 2 [11] and are compiled into a form that is enforced at run-time through a theorem prover. KAoS polices can be defined graphically however, unlike our approach, reasoning about KAoS models is performed symbolically. Both Rei [12] and the W3C P3P project [13] similarly allow providers of web services to model what personal information they use in a transaction and how that information is processed. The primary concern of Rei is that of specifying security control whereas P3P, in intent, extends the 'do not track' feature found in modern web browsers to include privacy options. Both the authorization mechanisms within Rei and the privacy mechanisms of P3P may be implemented by any software agent that interacts with the web. User interfaces for P3P present a number of controls to users who may customise their personal privacy policies. Again, like KAoS, Rei and P3P do not support diagrammatic reasoning. All reasoning is performed at a symbolic level.

In [14] the authors extend existing role-based access control (RBAC) mechanisms to support privacy. Their concepts of `OwnerConsent` and temporal periods closely reflect aspects of the 3CR model. This work similarly applies the *Purpose Specification Principle* and *Use Limitation Principal* from the OECD guidelines. Translating [14]'s RBAC policies into OWL can be achieved through either of the representations described in [15]. Our approach differs from [14] in the philosophical inspiration of the privacy model and in the number of transformations of the model in order to implement it. Our approach transforms an informal diagrammatic model into a formal diagrammatic model and from there allows implementation.

The diagrammatic modelling of privacy constraints in [16] allows, like our approach, diagrammatic reasoning to be performed with the constraints. The presentation in [16] is motivated by the need for clear communication of privacy legislation within large engineering teams. As such, UML diagrams – a notation favoured in real-world engineering – are the preferred diagrammatic notation. Through our use of concept-diagrams we provide an unambiguous model where the translation to implementation can be assisted using a theorem prover.

## 6 Conclusion

Diagrams in [3] were intended as illustrative rather than formal. We have demonstrated how an informal diagrammatic system can be reinterpreted as a formal concept diagram. Our reification resulted in a constraint-diagram that makes extensive use of axioms found in [4]. While we cannot offer proof that the formal system captures precisely the informal system, we have provided an implementation of the formal system. In our implemenation we rely on functionality provided by Haskell's software transactional memory. It is clear that our approach to implementing constraints may be extended to other transactional systems that allow expressive constraints.

In future work we wish to extend mechanisms for enforcing privacy in information systems. In particular we want to introduce notions of time into concept diagrams so that sharing of information may be bounded temporally. Moreover, we wish to algorithmically refine constraints in our concept diagram into suitable type-level and runtime transaction constraints.

## References

1. Warren, S.D., Brandeis, L.D.: The right to privacy. Harvard Law Rev. **4**(5), 193–220 (1890)
2. Floridi, L.: The ontological interpretation of informational privacy. Ethics Inf. Technol. **7**(4), 185–200 (2005)
3. Chen, S., Williams, M.: Privacy: an ontological problem. In: Pacific Asia Conference on Information Systems, PACIS, pp. 1402–1413 (2010)
4. Stapleton, G., Compton, M., Howse, J.: Visualizing OWL 2 using diagrams. In: Visual Languages and Human-Centric Computing. IEEE (2017)
5. Moor, J.H.: Towards a theory of privacy in the information age. SIGCAS Comput. Soc. **27**(3), 27–32 (1997)
6. Stapleton, G., Howse, J., Chapman, P., Delaney, A., Burton, J., Oliver, I.: Formalizing concept diagrams. In: Distributed Multimedia Systems, vol. 19, pp. 182–187. Knowledge Systems Institute (2013)
7. Howse, J., Stapleton, G., Taylor, J.: Spider diagrams. LMS J. Comput. Math. **8**, 145–194 (2005)
8. Chen, S., Williams, M.: Grounding data purpose and data usage for better privacy requirements development: an information system perspective. In: Pacific Asia Conference on Information Systems, PACIS (2011)
9. Harris, T., Jones, S.P.: Transactional memory with data invariants. In: SIGPLAN Workshop on Languages, Compilers, and Hardware Support for Transactional Computing. ACM (2006)
10. Bradshaw, J., Uszok, A., Jeffers, R., Suri, N., Hayes, P., Burstein, M., Acquisti, A., Benyo, B., Breedy, M., Carvalho, M., Diller, D.: Representation and reasoning for DAML-based policy and domain services in KAoS and Nomads. In: Joint Conference on Autonomous Agents and Multiagent Systems, pp. 835–842. ACM (2003)
11. Motik, B., Patel-Schneider, P.F., Parsia, B.: OWL 2 web ontology language: structural specification and functional-style syntax. http://www.w3.org/TR/owl2-syntax/. Accessed Dec 2017
12. Kagal, L., Finin, T., Paolucci, M., Srinivasan, N., Sycara, K., Denker, G.: Authorization and privacy for semantic web services. IEEE Intell. Syst. **19**(4), 50–56 (2004)
13. Cranor, L.F.: P3P: making privacy policies more useful. IEEE Secur. Privacy **99**(6), 50–55 (2003)
14. Ni, Q., Bertino, E., Lobo, J., Brodie, C., Karat, C.M., Karat, J., Trombeta, A.: Privacy-aware role-based access control. Trans. Inf. Syst. Secur. **13**(3), 24 (2010)
15. Finin, T., Joshi, A., Kagal, L., Niu, J., Sandhu, R., Winsborough, W., Thuraisingham, B.: ROWLBAC: representing role based access control in OWL. In: Access Control Models and Technologies, pp. 73–82. ACM (2008)
16. Oliver, I.: Privacy Engineering: A Dataflow and Ontological Approach. CreateSpace Independent Publishing Platform (2014)

# On Diagrams and General Model Checkers

Sheila R. M. Veloso[1(✉)], Paulo A. S. Veloso[2], Mario R. F. Benevides[2],
and Isaque M. S. Lima[2]

[1] Computing and System Engineering Department, Faculty of Engineering,
State University of Rio de Janeiro (UERJ), Rio de Janeiro, RJ, Brazil
sheila.murgel.bridge@gmail.com
[2] Computing and System Engineering Program, COPPE,
Institute of Mathematics, Federal University of Rio de Janeiro (UFRJ),
Rio de Janeiro, RJ, Brazil
pasveloso@gmail.com, {mario,isaque}@cos.ufrj.br

**Abstract.** We introduce a general approach, based on diagrams, to the specification and construction of model checkers. This approach gives general model checkers that can be instantiated to a model checker for a specific modal logic with semantics described by graphical rules. This paper proposes a way of combining graphical and general approaches to model checking so that the instantiation to specific logics is user-friendly and natural.

**Keywords:** Diagrams · Graphical approach
General model checkers · Graphical converters · Modalities

## 1 Introduction

We introduce a general approach, based on diagrams, to the specification and construction of model checkers. This approach gives general model checkers that can be instantiated to model checkers for specific logics with semantics described by graphical rules. We illustrate the approach by applying it to classical and intuitionistic modal logics.

Many new modal logics have been proposed for expressing properties of finite-state (concurrent) systems. Model checking is a very powerful and mature technique for this task. A model checker for a specific logic receives (representations of) a finite model $\mathfrak{M}$, a formula $\varphi$ and state $\mathsf{a}$ of $\mathfrak{M}$ and decides whether $\mathfrak{M}$ satisfies $\varphi$ at $\mathsf{a}$ [4].

Our aim is to provide a general graphical framework to express modal logics and to generate model checkers in an easy and neat way. Instead of developing a specific model checker for each modal logic, our approach will provide a single

---

Research partly sponsored by the Brazilian research agencies CNPq, FAPERJ and CAPES.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 680–688, 2018.
https://doi.org/10.1007/978-3-319-91376-6_60

general model checker that can be easily instantiated to one for each logic. Such a general model checker, which can be instantiated easily for a large class of logics, has several interesting possible applications in teaching and in investigating newly developed logics.

This approach involves two components: a logic-dependent converter and a general model checker; the former receives a formula $\varphi$ and converts it to a graphical expression E, whereas the latter, upon receiving E, a model $\mathfrak{M}$ and a state $\mathsf{a}$ of $\mathfrak{M}$, decides whether $\mathfrak{M}$ satisfies E at $\mathsf{a}$. The user will be concerned only with the graphical description of the semantics of the specific logic (see Fig. 1, p. 2). The converter will eliminate the symbols of the formula according to this description.

This approach has crucial issues concerning formulation of general rules and handling expressions; graphical concepts, based on diagrams, are very convenient for them. The visualization provided by diagrams is a useful guide for the translation of the semantics. The graphical machinery we employ here extends that used in [2] (for binary relations) by the addition of unary arcs (for subsets).

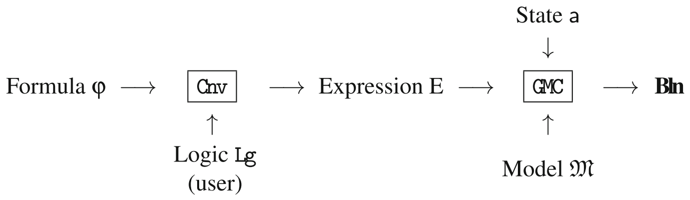


**Fig. 1.** General model checker GMC with converter Cnv

The structure of this paper is as follows. In Sect. 2, we introduce, informally, some basic ideas involved in graphical specification of logic. Sect. 3 presents the precise ideas underlying our graphical approach. In Sect. 4, we examine more precisely graphical formulations of semantics: graphical expressions and elimination rules. In Sect. 5, we introduce our general graphical converter and model checker and examine their instances. Sect. 6 presents some concluding remarks about our graphical approach.

## 2 Graphical Specification: Basic Ideas

We now introduce some basic ideas, which will be defined in Sect. 3.

Our graphical objects involve nodes and arcs. Nodes stand for states (see p. 4).

An important concept is satisfaction of a formula at a state; we use unary arcs for this. We represent that formula $\varphi$ holds at node w by a dashed line from w to $\varphi$: w – – –≺ $\varphi$ . We can also represent joint satisfaction at a node, e. g. p ≻– – – w – – –≺ q represents that both formulas p and q hold at node w.

Another important ingredient is state transition; we use binary arcs for this. We represent that node v is accessible from node u by the relation of r by a solid arrow labelled r from u to v: $u \xrightarrow{\mathsf{r}} v$ .

Sets of nodes and arcs will form drafts (see Sect. 3, p. 4). We now wish to represent that a node "sees" some node where a formula holds. For this purpose, we mark this node, obtaining a page (see Sect. 3, p. 4). Page $\widehat{u} \xrightarrow{\mathsf{r}} v - - -\!\!\prec p$ represents that some node r-accessible from node u satisfies formula p.

We also wish to represent non-satisfaction: we use complementation (noted by an overbar) for this. Page $\widehat{w} - - -\!\!\prec \overline{\varphi}$ represents that formula $\varphi$ does not hold at node w. We also consider complementation of pages obtaining expressions. Page $p \succ\!\!- - - \widehat{w} - - -\!\!\prec q$ represents that both formulas p and q hold at w, and expression $\overline{p \succ\!\!- - - \widehat{w} - - -\!\!\prec q}$ represents that formulas p and q do not hold simultaneously at w.

A (simple) *modal language* $\mathbb{ML}$ is characterized by 2 sets of symbols: $\mathbb{PL}$, of *propositional letters*, and $\mathbb{RN}$, of *relation names*. A model $\mathfrak{M}$ for $\mathbb{ML}$, over universe $\mathsf{M} \neq \emptyset$, assigns a subset $p^{\mathfrak{M}}$ of $\mathsf{M}$, to each propositional letter $p \in \mathbb{PL}$, and a 2-ary relation $\mathsf{r}^{\mathfrak{M}}$ on $\mathsf{M}$ to each relation name $\mathsf{r} \in \mathbb{RN}$.

We can now outline the semantics of our graphical concepts (see Sect. 3, p. 4). The behaviour of a page P is the set $[\![P]\!]_{\mathfrak{M}}$ consisting of the values assigned to its marked node by assignments satisfying its arcs. For a formula $\varphi$, we use $\varphi^{\mathfrak{M}}$ for the set of states satisfying $\varphi$ in $\mathfrak{M}$, and similarly $(E)_{\mathfrak{M}}$ for an expression E, with $(\overline{E})_{\mathfrak{M}} := \mathsf{M} \setminus (E)_{\mathfrak{M}}$.

This graphical machinery is already powerful enough to specify some fragments of classical modal logic $\mathbb{KM}$, as we now illustrate. We will first recall the notion of satisfaction of a formula at a state of a model and, then formulate it graphically.

We begin with the propositional fragment of classical modal logic $\mathbb{KM}$ having only negation $\neg$ and conjunction $\wedge$.

Satisfaction of formula $\varphi$ at state $\mathsf{a}$ of model $\mathfrak{C}$ is recursively defined [1]. In particular: $\mathfrak{C}, \mathsf{a} \Vdash \neg\varphi$ iff $\mathfrak{C}, \mathsf{a} \nVdash \varphi$ and $\mathfrak{C}, \mathsf{a} \Vdash \psi \wedge \theta$ iff $\mathfrak{C}, \mathsf{a} \Vdash \psi$ and $\mathfrak{C}, \mathsf{a} \Vdash \theta$.

We wish to formulate this semantics in a graphical manner. For this purpose, we provide appropriate graphical expressions: $\mathsf{a} \notin \varphi^{\mathfrak{C}}$ iff $\mathsf{a} \in (\overline{\varphi})_{\mathfrak{C}}$ and $\mathsf{a} \in \psi^{\mathfrak{C}}$ and $\mathsf{a} \in \theta^{\mathfrak{C}}$ iff $\mathsf{a} \in (\, \psi \succ\!\!- - - \widehat{x} - - -\!\!\prec \theta \,)_{\mathfrak{C}}$.

We thus obtain the graphical formulation in Example 1.

*Example 1 (Graphical classical propositional fragment semantics).* Satisfaction and expressions.

($\neg$) $\mathfrak{C}, \mathsf{a} \Vdash \neg\varphi$ iff $\mathfrak{C}, \mathsf{a} \nVdash \varphi$ iff $\mathsf{a} \notin \varphi^{\mathfrak{C}}$ iff $\mathsf{a} \in (\overline{\varphi})_{\mathfrak{C}}$ (complemented expression).

($\wedge$) $\mathfrak{C}, \mathsf{a} \Vdash \psi \wedge \theta$ iff $\mathfrak{C}, \mathsf{a} \Vdash \psi$ and $\mathfrak{C}, \mathsf{a} \Vdash \theta$ iff $\mathsf{a} \in \psi^{\mathfrak{C}}$ and $\mathsf{a} \in \theta^{\mathfrak{C}}$ iff $\mathsf{a} \in (\, \psi \succ\!\!- - - \widehat{x} - - -\!\!\prec \theta \,)_{\mathfrak{C}}$ (1-node page). ▹

We now consider a fragment of classical modal logic $\mathtt{KM}$ with negation $\neg$, conjunction $\wedge$ and modalities $\langle \mathsf{r} \rangle$. Negation $\neg$ and conjunction $\wedge$ are as before. So, we concentrate on the modalities $\langle \mathsf{r} \rangle$.

*Example 2 (Classical modal semantics).* Satisfaction and expressions.

1. Satisfaction for $\langle \mathsf{r} \rangle\, \varphi$: $\mathfrak{C}, \mathsf{a} \Vdash \langle \mathsf{r} \rangle\, \varphi$ iff for some $\mathsf{b}$ with $(\mathsf{a}, \mathsf{b}) \in \mathsf{r}^{\mathfrak{C}}$: $\mathfrak{C}, \mathsf{b} \Vdash \varphi$.
2. $\mathfrak{C}, \mathsf{a} \Vdash \langle \mathsf{r} \rangle\, \varphi$ iff for some $\mathsf{b}$ with $(\mathsf{a}, \mathsf{b}) \in \mathsf{r}^{\mathfrak{C}}$: $\mathfrak{C}, \mathsf{b} \Vdash \varphi$ iff for some $\mathsf{b}$ with $(\mathsf{a}, \mathsf{b}) \in \mathsf{r}^{\mathfrak{C}}$: $\mathsf{b} \in \varphi^{\mathfrak{P}}$ iff $\mathsf{a} \in (\, \hat{x} \xrightarrow{\mathsf{r}} y \mathrel{-\,-\,-\!\!\prec} \varphi \,)_{\mathfrak{C}}$. (2-node page). $\flat$

We now consider another graphical concept. A book is a finite set of (alternative) pages (see Sect. 3, p. 4). For instance, the empty book $\{\ \}$ has no pages. The behaviour of a book is the union of the behaviours of its pages, e. g. $\llbracket \{\ \} \rrbracket_{\mathfrak{M}} = \emptyset$ (see Sect. 3, p. 4).

We now can extend our fragment of classical modal logic $\mathtt{KM}$ by adding absurdity $\bot$ and disjunction $\vee$. We will focus on the new symbols.

*Example 3 (Extended classical modal semantics).* Satisfaction and expressions.

($\bot$) Absurdity $\bot$: $\mathfrak{C}, \mathsf{a} \nVdash \bot$, i. e. $\bot^{\mathfrak{C}} = \emptyset$; so $\bot^{\mathfrak{C}} = \emptyset = \llbracket \{\ \} \rrbracket_{\mathfrak{C}}$ (empty book).

($\vee$) Disjunction $\vee$: $\mathfrak{C}, \mathsf{a} \Vdash \psi \vee \theta$ iff $\mathfrak{C}, \mathsf{a} \Vdash \psi$ or $\mathfrak{C}, \mathsf{a} \Vdash \theta$, i. e. $(\psi \vee \theta)^{\mathfrak{C}} = \psi^{\mathfrak{C}} \cup \theta^{\mathfrak{C}}$; so $(\psi \vee \theta)^{\mathfrak{C}} = \psi^{\mathfrak{C}} \cup \theta^{\mathfrak{C}} = \Big( \{\ \psi \mathrel{\succ\!\!-\,-\,-} \hat{x}\ ,\ \hat{x} \mathrel{-\,-\,-\!\!\prec} \theta\ \} \Big)_{\mathfrak{C}}$ (2-page book). $\flat$

*Example 4 (Classical modal defined symbols).* In classical modal logic $\mathtt{KM}$, conditional and the box modalities can be introduced by definitions, namely: $\psi \rightarrow \theta := (\neg\psi) \vee \theta$ and $[\mathsf{r}]\, \varphi := \neg \langle \mathsf{r} \rangle \neg\, \varphi$. With these definitions, Examples 1 and 3 (p. 4) give immediately the following graphical translations.

($\rightarrow$) $\mathfrak{C}, \mathsf{a} \Vdash \psi \rightarrow \theta$ iff $\mathsf{a} \in \Big( \underline{\{\ \overline{\psi} \mathrel{\succ\!\!-\,-\,-} \hat{x}\ ,\ \hat{x} \mathrel{-\,-\,-\!\!\prec} \theta\ \}} \Big)_{\mathfrak{C}}$ (2-page book).

($[\,]$) $\mathfrak{C}, \mathsf{a} \Vdash [\mathsf{r}]\, \varphi$ iff $\mathsf{a} \in (\, \hat{x} \xrightarrow{\mathsf{r}} y \mathrel{-\,-\,-\!\!\prec} \overline{\varphi} \,)_{\mathfrak{C}}$ (complemented 2-node page). $\flat$

## 3 Graphical Concepts and Constructions

We now introduce graphical concepts. For more details see, e. g. [2,7].

(E) The *expressions* are the formulas, the pages, the books and their complements (represented by an overbar); they will represent sets of states.

(a) Arcs are 1-ary or 2-ary. *Unary arc*: $w \mathrel{-\,-\,-\!\!\prec} E$ (node $w$, expression $E$). *Binary arc*: $u \xrightarrow{\mathsf{r}} v$ (nodes $u, v$, relation name $\mathsf{r}$).

- ($\Delta$) A *draft* consists of finite sets N, of nodes, and $\mathtt{A}$, of arcs. An example of a 4-node draft is $\Delta \;=\; p \succ\!- - - u \xrightarrow{\mathsf{r}} v \qquad w \xrightarrow{\mathsf{s}} z - - -\!\!\prec q$ .
- (P) A *page* P consists of an *underlying draft* $\underline{\mathrm{P}}$ together with a *link node* (marked $\widehat{\ }$). For instance: $\mathrm{P} \;=\; p \succ\!- - - \widehat{u} \xrightarrow{\mathsf{r}} v \qquad w \xrightarrow{\mathsf{s}} z - - -\!\!\prec q$ is a page.
- (B) A *book* is a finite set of pages. An example of a 2-page book is $\{\,\mathrm{P}, \widehat{z}\,\}$.

Graphical semantics is as follows. Consider a model $\mathfrak{M}$, over universe $\mathsf{M}$ (cf. p. 3).

- (E) *Set of expression*: for a formula $\varphi$, $(\varphi)_{\mathfrak{M}} := \varphi^{\mathfrak{M}}$, if expression E is a page or a book, then $(\mathrm{E})_{\mathfrak{M}} := [\![\mathrm{E}]\!]_{\mathfrak{M}}$ (see below); $(\overline{\mathrm{E}})_{\mathfrak{M}} := \mathsf{M} \setminus (\mathrm{E})_{\mathfrak{M}}$.
- ($\Vdash$) *Satisfaction* under assignment $\mathsf{g} : \mathrm{N} \to \mathsf{M}$ (assigning to node $w \in \mathrm{N}$ state $w^{\mathsf{g}} \in \mathsf{M}$).
  - ($\mathsf{a}$) $\mathfrak{M}, \mathsf{g} \Vdash\; w - - -\!\!\prec \mathrm{E}$ iff $w^{\mathsf{g}} \in (\mathrm{E})_{\mathfrak{M}}$ and $\mathfrak{M}, \mathsf{g} \Vdash\; u \xrightarrow{\mathsf{r}} v$ iff $(u^{\mathsf{g}}, v^{\mathsf{g}}) \in \mathsf{r}^{\mathfrak{M}}$.
  - ($\Delta$) Draft: $\mathfrak{M}, \mathsf{g} \Vdash \Delta$ iff $\mathfrak{M}, \mathsf{g} \Vdash \mathsf{a}$, for every arc $\mathsf{a}$ of $\Delta$.
- (Pg) The *behaviour of page* P, with link w and underlying draft $\underline{\mathrm{P}}$, is the set of all values $w^{\mathsf{g}} \in \mathsf{M}$ for the assignments $\mathsf{g}$ satisfying $\underline{\mathrm{P}}$, i. e. $[\![\mathrm{P}]\!]_{\mathfrak{M}} := \{w^{\mathsf{g}} \in \mathsf{M} \,/\, \mathfrak{M}, \mathsf{g} \Vdash \underline{\mathrm{P}}\}$.
- (Bk) *Behaviour of book*: $[\![\mathsf{B}]\!]_{\mathfrak{M}} := \bigcup_{\mathrm{P} \in \mathsf{B}} [\![\mathrm{P}]\!]_{\mathfrak{M}}$.
- ($\equiv$) Expressions E and F are *equivalent* in a class $\mathbb{K}$ of models iff $(\mathrm{E})_{\mathfrak{M}} = (\mathrm{F})_{\mathfrak{M}}$, for every model $\mathfrak{M} \in \mathbb{K}$.

We introduce graphical constructions for expression set $\boldsymbol{F} = \{\,\mathrm{F}_1, \ldots, \mathrm{F}_h\,\}$ and expression E. They aim at capturing simultaneous and alternative satisfaction, as well as change under transition (see Lemma 1: Graphical constructions, p. 5).

(PG) *Page of expression set* $\mathrm{PG}(\boldsymbol{F}) :=$
$$\begin{tikzcd} \widehat{x} \arrow[r, dashed, "\prec" very near end] \arrow[rdd, dashed, "\prec" very near end] & \mathrm{F}_1 \\ & \vdots \\ & \mathrm{F}_h \end{tikzcd}$$
$\mathrm{Pg}(\mathrm{E}) := \mathrm{PG}(\{\mathrm{E}\})$.

(BK) *Book of expression set*:
$$\mathrm{BK}[\boldsymbol{F}] := \{\mathrm{Pg}(\mathrm{F}_1), \ldots, \mathrm{Pg}(\mathrm{F}_h)\} = \{\, \widehat{x} - - -\!\!\prec \mathrm{F}_1 \,, \ldots, \; \widehat{x} - - -\!\!\prec \mathrm{F}_h \,\}.$$

(FP) *Follow-page* (for $\mathsf{r} \in \mathrm{RN}$) $\mathrm{FP}(\mathsf{r}, \boldsymbol{F}) :=$
$$\begin{tikzcd} \widehat{x} \arrow[r, "\mathsf{r}"] & y \arrow[r, dashed, "\prec" very near end] \arrow[rdd, dashed, "\prec" very near end] & \mathrm{F}_1 \,. \\ & & \vdots \\ & & \mathrm{F}_h \end{tikzcd}$$

We use the abbreviation: $\mathrm{fp}(\mathsf{r}, \mathrm{E}) := \mathrm{FP}(\mathsf{r}, \{\mathrm{E}\}) \;=\; \widehat{x} \xrightarrow{\mathsf{r}} y - - -\!\!\prec \mathrm{E}$ .

These concepts will be used for formulating semantical clauses in the sequel.

**Lemma 1 (Graphical constructions).** $[\![\mathrm{Pg}(\mathrm{E})]\!]_{\mathfrak{M}} = (\mathrm{E})_{\mathfrak{M}}$, $[\![\mathrm{fp}(\mathsf{r}, \mathrm{E})]\!]_{\mathfrak{M}} = \{\mathsf{a} \in \mathsf{M} \,/\, \exists\, \mathsf{b} \in \mathsf{M}\, ((\mathsf{a}, \mathsf{b}) \in \mathsf{r}^{\mathfrak{M}} \wedge \mathsf{b} \in (\mathrm{E})_{\mathfrak{M}})\}$*;* $[\![\mathrm{PG}(\boldsymbol{F})]\!]_{\mathfrak{M}} = \bigcap_{\mathrm{F} \in \boldsymbol{F}} (\mathrm{F})_{\mathfrak{M}}$, $[\![\mathrm{BK}[\boldsymbol{F}]]\!]_{\mathfrak{M}} = \bigcup_{\mathrm{F} \in \boldsymbol{F}} (\mathrm{F})_{\mathfrak{M}}$ *and* $[\![\mathrm{FP}(\mathsf{r}, \boldsymbol{F})]\!]_{\mathfrak{M}} = \{\mathsf{a} \in \mathsf{M} \,/\, \exists\, \mathsf{b} \in \mathsf{M}\, ((\mathsf{a}, \mathsf{b}) \in \mathsf{r}^{\mathfrak{M}} \wedge \mathsf{b} \in [\![\mathrm{PG}(\boldsymbol{F})]\!]_{\mathfrak{M}})\}$. $\natural$

*Proof.* The assertions follow from graphical semantics (p. 4).

We also use *neat assertions* and finite *neat-assertion sets*, jointly generated by:

$$\begin{array}{ll} \mathrm{F} ::= \mathrm{p} \mid \overline{\mathrm{F}} \mid \mathbb{PG}(\boldsymbol{F}) \mid \mathsf{BK}[\boldsymbol{F}] \mid \mathbb{FP}(\mathsf{r}, \boldsymbol{F}) & (\mathrm{p} \in \mathbb{PL}, \mathsf{r} \in \mathbb{RN}) \\ \boldsymbol{F} ::= \emptyset \mid \boldsymbol{F} \cup \{\mathrm{F}\} & (\boldsymbol{F} = \{\mathrm{F}_1, \ldots, \mathrm{F}_h\}) \end{array} \quad (1)$$

## 4 Graphical Formulations of Semantics

A *(graphical) specification* for a logic $\mathbb{Lg}$ consists of rules $\varphi := \mathrm{E}$, with formula $\varphi$ and expression E. In Sect. 2 (Graphical Specification: Basic Ideas), we have indicated how one can obtain a specification for the Kripke semantics of classical modal logic $\mathbb{KM}$ (cf. Examples 1, p. 3, to 4, p. 4). Figure 2 (p. 5) summarizes this specification with our graphical constructs (cf. Sect. 3: Graphical Concepts and Constructions, p. 5.)

| Formula: | $\bot$ | $\neg\varphi$ | $\psi \wedge \theta$ | $\psi \vee \theta$ | $\psi \rightarrow \theta$ | $\langle \mathsf{r} \rangle \varphi$ | $[\mathsf{r}]\varphi$ |
|---|---|---|---|---|---|---|---|
| Construct: | $\mathsf{BK}[\emptyset]$ | $\overline{\varphi}$ | $\mathbb{PG}(\{\psi, \theta\})$ | $\mathsf{BK}[\{\psi, \theta\}]$ | $\mathsf{BK}[\{\overline{\psi}, \theta\}]$ | $\mathsf{fp}(\mathsf{r}, \varphi)$ | $\overline{\mathsf{fp}(\mathsf{r}, \overline{\varphi})}$ |

**Fig. 2.** Graphical constructs for classical modal logic $\mathbb{KM}$

Thus, we obtain the *elimination rules* for $\mathbb{KM}$ shown in Fig. 3 (p. 6).

$$\mathbb{ER}_{\mathbb{KM}}(\bot) := \mathsf{BK}[\emptyset] \quad \mathbb{ER}_{\mathbb{KM}}(\neg)[\varphi] := \overline{\varphi} \quad \mathbb{ER}_{\mathbb{KM}}(\wedge)[\psi, \theta] := \mathbb{PG}(\{\psi, \theta\}) \quad \mathbb{ER}_{\mathbb{KM}}(\vee)[\psi, \theta] := \mathsf{BK}[\{\psi, \theta\}]$$

$$\mathbb{ER}_{\mathbb{KM}}(\rightarrow)[\psi, \theta] := \mathsf{BK}[\{\overline{\psi}, \theta\}] \quad \mathbb{ER}_{\mathbb{KM}}(\langle \mathsf{r} \rangle)[\varphi] := \mathsf{fp}(\mathsf{r}, \varphi) \quad \mathbb{ER}_{\mathbb{KM}}([\mathsf{r}])[\varphi] := \overline{\mathsf{fp}(\mathsf{r}, \overline{\varphi})}$$

**Fig. 3.** Elimination rules for classical modal logic $\mathbb{KM}$

We now consider intuitionistic modal logic $\mathbb{JM}$ [5].

*Example 5 (Intuitionistic modal semantics).* Consider an intuitionistic modal model $\mathfrak{I}$ with world precedence $\preceq^{\mathfrak{I}}$.

1. Satisfaction of formula $\varphi$ at state $\mathsf{a}$ of $\mathfrak{I}$ is recursively defined as follows [5]. For $\mathrm{p}, \bot, \wedge, \vee$, and $\langle\,\rangle$: as in classical modal semantics (cf. Sect. 2, p. 3). For $\neg$: $\mathfrak{I}, \mathsf{a} \Vdash \neg\varphi$ iff for every $\mathsf{b}$ with $(\mathsf{a}, \mathsf{b}) \in \preceq^{\mathfrak{I}}$: $\mathfrak{I}, \mathsf{b} \nVdash \varphi$ (i. e. there is no $\mathsf{b}$ such that $(\mathsf{a}, \mathsf{b}) \in \preceq^{\mathfrak{I}}$ and $\mathfrak{I}, \mathsf{b} \Vdash \varphi$). For $\rightarrow$: $\mathfrak{I}, \mathsf{a} \Vdash \psi \rightarrow \theta$ iff whenever $(\mathsf{a}, \mathsf{b}) \in \preceq^{\mathfrak{I}}$: if $\mathfrak{I}, \mathsf{b} \Vdash \psi$ then $\mathfrak{I}, \mathsf{b} \Vdash \theta$ (i. e. there is no $\mathsf{b}$ with $(\mathsf{a}, \mathsf{b}) \in \preceq^{\mathfrak{I}}$ such that $\mathfrak{I}, \mathsf{b} \Vdash \psi$ and $\mathfrak{I}, \mathsf{b} \nVdash \theta$). For $[\mathsf{r}]$: $\mathfrak{I}, \mathsf{a} \Vdash [\mathsf{r}]\varphi$ iff whenever $(\mathsf{a}, \mathsf{b}) \in \preceq^{\mathfrak{I}}$ and $(\mathsf{b}, \mathsf{c}) \in \mathsf{r}^{\mathfrak{I}}$: $\mathfrak{I}, \mathsf{c} \Vdash \varphi$ (i. e. there are no $\mathsf{b}, \mathsf{c}$ with $(\mathsf{a}, \mathsf{b}) \in \preceq^{\mathfrak{I}}$, $(\mathsf{b}, \mathsf{c}) \in \mathsf{r}^{\mathfrak{I}}$ and $\mathfrak{I}, \mathsf{c} \nVdash \varphi$).

2. We now provide the graphical formulation.

(¬) Formula $\neg\,\varphi$ $\mathbb{JM}$-equivalent to $\overline{\hat{x} \xrightarrow{\preceq} y \;{-}\,{-}\,{-}\!\!\prec \varphi}$ .

(→) Formula $\psi \to \theta$ $\mathbb{JM}$-equivalent to

$$\begin{tikzcd} \hat{x} \arrow[r, "\preceq"] & y \arrow[r, dashed, "\prec"] \arrow[dr, dashed, "\prec"'] & \psi \, . \\ & & \overline{\theta} \end{tikzcd}$$

([ ]) Formula $[\mathsf{r}]\,\varphi$ $\mathbb{JM}$-equivalent to $\overline{\hat{x} \xrightarrow{\preceq} y \xrightarrow{\mathsf{r}} z \;{-}\,{-}\,{-}\!\!\prec \overline{\varphi}}$ .

(⋆) For $\bot, \wedge, \vee$ and $\langle \mathsf{r} \rangle$, as in classical modal semantics (cf. Sct. 2, p. 3). ♭

Figure 4 (p. 6) shows the *elimination rules* for the intuitionistic connectives ¬ and → and for the intuitionistic modality $[\mathsf{r}]$.

$$\mathbb{ER}_{\mathbb{JM}}(\neg)[\varphi] := \overline{\mathsf{fp}(\preceq, \varphi)} \quad \mathbb{ER}_{\mathbb{JM}}(\rightarrow)[\psi, \theta] := \overline{\mathsf{fp}(\preceq, \mathbb{PG}(\{\psi, \overline{\theta}\}))} \quad \mathbb{ER}_{\mathbb{JM}}([\mathsf{r}])[\varphi] := \overline{\mathsf{fp}(\preceq, (\mathsf{fp}(\mathsf{r}, \overline{\varphi})))}$$

**Fig. 4.** Elimination rules for intuitionistic ¬, → and $[\mathsf{r}]$

**Proposition 1 (Graphical specifications).** *Consider the specifications for* $\mathbb{KM}$ *(cf. Fig. 3, p. 6) and for* $\mathbb{JM}$ *(cf. Fig. 4, p. 6). For every elimination rule* $\varphi := \mathrm{E}$*: formula* $\varphi$ *is equivalent to expression* $\mathrm{E}$ *in the corresponding class of models.* ♮

*Proof.* By the graphical constructs (cf. Sect. 3, p. 5) and each semantics.

In general, graphical formulation of semantics (as illustrated in the preceding examples) involves two steps as follows.

1. Translate the satisfaction clauses to graphical expressions. Cf. Examples 1 to 3 (p. 3) and 5 (p. 6). Here, the visualization provided by diagrams is very useful.
2. Formulate the resulting expressions by means of expression constructs $\mathbb{PG}$, $\mathbb{BK}$ and $\mathbb{FP}$ (cf. Sect. 3: Graphical Concepts and Constructions, p. 5).

## 5 General Converter and Model-Checker

We will consider modal languages, with a set $\mathbb{PL}$ of propositional letters, each one characterized by its 0-ary, 1-ary and 2-ary connectives (†, $\nabla$ and $\bullet$) and modalities ($\mu \in \Xi$). Such a language $\mathbb{ML}$ has set $\Phi$ of *formulas* generated by the grammar:

$$\varphi \ ::= \ \mathrm{p} \mid \dagger \mid \nabla\,\varphi \mid \varphi' \bullet \varphi'' \mid \mu\,\varphi \qquad (\mathrm{p} \in \mathbb{PL},\ \mu \in \Xi) \tag{2}$$

As a special case, we have the language $\mathbb{ML}$ for $\mathbb{KM}$ and $\mathbb{JM}$ (cf. Sect. 2, p. 3).

A general converter generates, for each logic $\mathbb{Lg}$, a converter $\mathbb{GC}_{\mathbb{Lg}}$, which eliminates logical symbols producing neat expressions (cf. grammar (1), p. 5). We define our *general converter* $\mathbb{GC}$ : Logic $\to$ (Formulas $\to$ Expressions) as follows:

$$\mathbb{GC}_{\mathbb{Lg}}(\mathrm{p}) := \mathrm{p}; \qquad \mathbb{GC}_{\mathbb{Lg}}(\dagger) := \mathbb{ER}_{\mathbb{Lg}}(\dagger); \qquad \mathbb{GC}_{\mathbb{Lg}}(\nabla\,\varphi) := \mathbb{ER}_{\mathbb{Lg}}(\nabla)[\mathbb{GC}_{\mathbb{Lg}}(\varphi)];$$
$$\mathbb{GC}_{\mathbb{Lg}}(\psi\ \bullet\ \theta) := \mathbb{ER}_{\mathbb{Lg}}(\bullet)[\mathbb{GC}_{\mathbb{Lg}}(\psi), \mathbb{GC}_{\mathbb{Lg}}(\theta)];\ \mathbb{GC}_{\mathbb{Lg}}(\mu\,\varphi) := \mathbb{ER}_{\mathbb{Lg}}(\mu)[\mathbb{GC}_{\mathbb{Lg}}(\varphi)].$$

We now consider instantiated graphical converters for $\mathbb{KM}$ and $\mathbb{JM}$: from formulas to neat expressions. It suffices to apply their elimination rules: e. g. $\mathbb{GC}_{\mathbb{KM}}(\langle \mathsf{r}\rangle\,[\mathsf{s}]\,\neg\,\mathrm{p}) = \mathsf{fp}(\mathsf{r}\,,\,(\overline{\mathsf{fp}(\mathsf{s}\,,\,\overline{\overline{\mathrm{p}}})}))$ and $\mathbb{GC}_{\mathbb{JM}}(\langle \mathsf{r}\rangle\,[\mathsf{s}]\,\neg\,\mathrm{p}) = \mathsf{fp}(\mathsf{r}\,,\,(\mathsf{fp}(\preceq\,,\,(\mathsf{fp}(\mathsf{s}\,,\,\overline{\overline{\mathsf{fp}(\preceq\,,\,\mathrm{p})}}))))))$.

We now consider our general graphical model checker.

We will use a Boolean sort **Bln**, having 2 values **false** and **true**, with the usual 2-ary operations **<u>and</u>** and **<u>or</u>**, as well as the 1-ary operation **<u>not</u>**, which we extend naturally to finite sets. We will also use *reachable sets* consisting of the states reached under a transition: $[\mathsf{a}^{\mathsf{r}}]_{\mathfrak{M}} := \{\mathsf{b} \in \mathsf{M} \,/\, (\mathsf{a},\mathsf{b}) \in \mathsf{r}^{\mathfrak{M}}\}$, for a relation name $\mathsf{r} \in \mathbb{RN}$ (cf. Sect. 2, p. 3).

A general model checker receives a finite model $\mathfrak{M}$, a state $\mathsf{a}$ of $\mathfrak{M}$ and a neat expression E (cf. grammar (1), p. 5), and checks whether $\mathfrak{M}$ satisfies E at $\mathsf{a}$. Our *general graphical model checker* is a function $\mathbb{GMc}$ : ( Model $\mathfrak{M}$ , State $\mathsf{a}$ : Neat expression E ) $\to$ **Bln**. We use simply $\mathbb{GMc}(\mathsf{a} : \mathrm{E})$, when model $\mathfrak{M}$ is clear. We define $\mathbb{GMc}(\mathsf{a} : \mathrm{E})$ based on the form of E as follows: for a propositional letter $\mathrm{p} \in \mathbb{PL}$, $\mathbb{GMc}(\mathsf{a} : \mathrm{p}) := \mathbf{true}$ iff $\mathsf{a} \in \mathrm{p}^{\mathfrak{M}}$. $\mathbb{GMc}(\mathsf{a} : \overline{\mathrm{F}}) := \underline{\mathbf{not}}\,\mathbb{GMc}(\mathsf{a} : \mathrm{F})$; $\mathbb{GMc}(\mathsf{a} : \mathrm{PG}(\boldsymbol{F})) := \underline{\mathbf{and}}(\{\mathbb{GMc}(\mathsf{a} : \mathrm{F}) \,/\, \mathrm{F} \in \boldsymbol{F}\})$; $\mathbb{GMc}(\mathsf{a} : \mathbb{BK}[\boldsymbol{F}]) := \underline{\mathbf{or}}[\{\mathbb{GMc}(\mathsf{a} : \mathrm{F}) \,/\, \mathrm{F} \in \boldsymbol{F}\}]$; $\mathbb{GMc}(\mathsf{a} : \mathbb{FP}(\mathsf{r}\,,\boldsymbol{F})) := \underline{\mathbf{or}}[\{\mathbb{GMc}(\mathsf{b} : \mathrm{PG}(\boldsymbol{F})) \,/\, \mathsf{b} \in [\mathsf{a}^{\mathsf{r}}]_{\mathfrak{M}}\}]$.

We can instantiate our general graphical model checker $\mathbb{GMc}$ by a logic $\mathbb{Lg}$ (cf. Fig. 1, p. 2). We obtain a *graphical model checker for logic* $\mathbb{Lg}$, defined as follows: $\mathbb{GMc}_{\mathbb{Lg}}(\mathfrak{M}, \mathsf{a} : \varphi) := \mathbb{GMc}(\,\mathfrak{M}, \mathsf{a} :\, \mathbb{GC}_{\mathbb{Lg}}(\varphi)\,)$.

*Example 6 (Graphical model checker for* $\mathbb{KM}$*).* Consider a model $\mathfrak{C}$ for $\mathbb{KM}$ with universe $\mathsf{M} = \{\mathsf{a},\mathsf{b},\mathsf{c},\mathsf{d}\}$, subsets $\mathrm{p}^{\mathfrak{C}} = \{\mathsf{d}\}$ and $\mathrm{q}^{\mathfrak{C}} = \{\mathsf{b}\}$, of $\mathsf{M}$, and 2-ary relations $\mathsf{r}^{\mathfrak{C}} = \{(\mathsf{a},\mathsf{b}),(\mathsf{a},\mathsf{c})\}$ and $\mathsf{s}^{\mathfrak{C}} = \{(\mathsf{c},\mathsf{d})\}$ on $\mathsf{M}$. We wish to check whether formula $\varphi = \langle \mathsf{r}\rangle\,[\mathsf{s}]\,\neg\,\mathrm{p}$ holds at state $\mathsf{a}$ of model $\mathfrak{C}$. Model checker $\mathbb{GMc}(\mathsf{a} : \varphi)$, short for $\mathbb{GMc}_{\mathbb{KM}}(\mathfrak{C}, \mathsf{a} : \varphi)$, operates as follows: $\mathbb{GMc}(\mathsf{a} : \langle \mathsf{r}\rangle\,[\mathsf{s}]\,\neg\,\mathrm{p}) = \mathbb{GMc}(\mathsf{a} : \mathsf{fp}(\mathsf{r}\,,\,(\overline{\mathsf{fp}(\mathsf{s}\,,\,\overline{\overline{\mathrm{p}}})}))) = \mathbb{GMc}(\mathsf{b} : \overline{\mathsf{fp}(\mathsf{s}\,,\,\overline{\overline{\mathrm{p}}})})\,\underline{\mathbf{or}}\,\mathbb{GMc}(\mathsf{c} : \overline{\mathsf{fp}(\mathsf{s}\,,\,\overline{\overline{\mathrm{p}}})}) = \underline{\mathbf{not}}\,\mathbb{GMc}(\mathsf{b} : \mathsf{fp}(\mathsf{s}\,,\,\overline{\overline{\mathrm{p}}}))\,\underline{\mathbf{or}}\,\underline{\mathbf{not}}\,\mathbb{GMc}(\mathsf{c} : \mathsf{fp}(\mathsf{s}\,,\,\overline{\overline{\mathrm{p}}})) = \underline{\mathbf{not}}\,\mathbf{false}\,\underline{\mathbf{or}}\,\underline{\mathbf{not}}\,\mathbb{GMc}(\mathsf{d} : \overline{\overline{\mathrm{p}}}) = \mathbf{true}\,\underline{\mathbf{or}}\,\underline{\mathbf{not}}\,\mathbb{GMc}(\mathsf{d} : \mathrm{p}) = \mathbf{true}\,\underline{\mathbf{or}}\,\underline{\mathbf{not}}\,\mathbf{true} = \mathbf{true}\,\underline{\mathbf{or}}\,\mathbf{false} = \mathbf{true}$. $\flat$

*Example 7 (Graphical model checker for* $\mathbb{JM}$*).* Consider model $\mathfrak{J}$ for $\mathbb{JM}$ with universe $\mathsf{M} = \{\mathsf{a},\mathsf{b},\mathsf{c},\mathsf{d},\mathsf{e},\mathsf{f}\}$; subsets $\mathrm{p}^{\mathfrak{J}} = \{\mathsf{f}\}$ and $\mathrm{q}^{\mathfrak{J}} = \{\mathsf{b},\mathsf{d}\}$; and relations $\mathsf{r}^{\mathfrak{J}} = \{(\mathsf{a},\mathsf{b}),(\mathsf{a},\mathsf{c}),(\mathsf{a},\mathsf{d}),(\mathsf{a},\mathsf{e})\}$, $\mathsf{s}^{\mathfrak{J}} = \{(\mathsf{c},\mathsf{f}),(\mathsf{e},\mathsf{f})\}$ and $\prec^{\mathfrak{J}} = \{(\mathsf{b},\mathsf{d}),(\mathsf{c},\mathsf{e})\}$ (we use $\prec$ for ‘strictly precedes’: “precedes and different”). We wish to check whether formula $\varphi = \langle \mathsf{r}\rangle\,[\mathsf{s}]\,\neg\,\mathrm{p}$ holds at state $\mathsf{a}$ of model $\mathfrak{J}$. Model checker $\mathbb{GMc}(\mathsf{a} : \varphi)$, short for $\mathbb{GMc}_{\mathbb{JM}}(\mathfrak{J}, \mathsf{a} : \varphi)$, operates as follows: $\mathbb{GMc}(\mathsf{a} : \varphi) = \mathbb{GMc}(\mathsf{a} : \mathsf{fp}(\mathsf{r}, (\mathsf{fp}(\preceq, (\mathsf{fp}(\mathsf{s}, \overline{\overline{\mathsf{fp}(\preceq,\ \mathrm{p})}})))))) = \mathbf{true}$. $\flat$

## 6 Concluding Remarks

We have introduced a general graphical approach to specification and construction of model checkers, which gives general model checkers that can be instantiated to model checkers for specific logics. Our approach involves two components: a logic-dependent converter and a general model checker; the user is concerned only with the former (cf. Fig. 1, p. 2), which can be obtained by instantiating a general converter. We have illustrated the application of this approach to two modal logics: classical $\mathbb{KM}$ and intuitionistic $\mathbb{JM}$ (which clearly applies to their propositional fragments).

Our approach is rather flexible. It can be easily used to handle some relational constants and special modalities, like the universal operator $\mathsf{E}$ and the difference operator $\mathsf{D}$ [1]. Also, it can be extended to to structured modalities (as in PDL [3]): it suffices to describe the semantics of the structured relations (see, e. g. [6]). Tests cause no problem: we can eliminate $\langle\psi?\rangle\,\theta$ and $[\psi?]\,\theta$ to the 1-node page $\mathbb{PG}(\{\psi, \theta\})$.

We finally comment on the user's task: graphical specification of semantics. As illustrated in Sect. 4 this is not too difficult. One often reasons intuitively about modal semantics also in a graphical manner: our approach formalizes such intuitive arguments. We think we have combined graphical and general approaches in an interesting way.

The main feature of our approach is its generality, which provides flexibility without loss of efficiency. A user-friendly way of obtaining an instantiation adequate to a particular logic is crucial for the general applicability of this method. We think that the proposed graphical notation for the satisfaction conditions and its processing by a uniform method provide an interesting approach to these issues.

## References

1. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge University Press, Cambridge (2001)
2. de Freitas, R., Veloso, P.A.S., Veloso, S.R.M., Viana, P.: A calculus for graphs with complement. In: Goel, A.K., Jamnik, M., Narayanan, N.H. (eds.) Diagrams 2010. LNCS (LNAI), vol. 6170, pp. 84–98. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14600-8_11
3. Harel, D., Kozen, D., Tiuryn, J.: Dynamic Logic. MIT Press, Cambridge (2000)
4. Peled, D.A., Clarke, E.M., Grumberg, O.: Model Checking. MIT Press, Cambridge (2000)
5. Simpson, A.: The proof theory and semantics of intuitionistic modal logic. Ph.D. thesis, Edinburgh University, Edinburgh (1994)
6. Veloso, P., Veloso, S., Benevides, M.: PDL for structured data: a graph-calculus approach. Logic J. IGPL **22**, 737–757 (2014)
7. Veloso, P., Veloso, S., Benevides, M.: On graph calculi for multi-modal logics. In: Ayala-Rincón, M., Mackie, I. (eds.) Electronic Notes in Theoretical Computer Science, ENTCS, vol. 352, pp. 231–252 (2015)

# Depicting the Redundancy of Fourth Figure Using Venn-Peirce Framework

Sumanta Sarathi Sharma(✉)

School of Philosophy and Culture, Shri Mata Vaishno Devi University,
Katra Reasi 182 320, Jammu and Kashmir, India
ss.sharma@smvdu.ac.in

**Abstract.** An incessant debate that history of syllogistic reasoning has witnessed is on the status of fourth figure after its alleged invention. Commentators on Aristotle and several other logicians have advocated various approaches to include or abandon this last figure. However, in the middle of last century, the debate seemed to have reached quiescence with fifteen valid syllogisms present in four figures. Among this, some moods are distinct, i.e. they are valid in one figure whereas others are non-distinct as they are valid in multiple figures. In this paper, the notion of diagrammatic congruence for non-distinct syllogisms using Venn-Peirce diagrams is introduced. Consequently, we establish the equivalence of moods that are diagrammatically congruent. Furthermore, it is argued that the presence of a distinct mood is pivotal to recognize an arrangement as a separate figure, which is evident in Aristotle's own treatment of figures. With this, the redundancy of fourth figure is demonstrated.

**Keywords:** Fourth figure · Galen · Moods · Syllogistic reasoning
Venn-Peirce Diagrams

## 1 Introduction

Aristotle recognized three figures and found fourteen syllogisms valid. The number of syllogisms that are valid rose to twenty four (pooling the strengthened and weakened moods) with the inclusion of the fourth figure. The inclusion of this last figure to syllogistic reasoning has been a matter of controversy, ever since its inception. The disagreement pertaining to this is on two grounds. First, who has invented the fourth figure? And the second question, does this figure encapsulate the essence of Aristotle's syllogistic?

Galen is usually accredited with the discovery of fourth figure after Averroes' account [7]. However, it was always surrounded by allegations and apprehensions. The invention of fourth figure may be attributed to an unknown scholar, [9] as it is certain now that at least Galen has not invented it [14,17]. Moreover, this question can be done away with, as it is a minor contribution to syllogistic [14].

The second question on the relevance of fourth figure, however, is far more significant and a formidable one to tackle. The present paper addresses this question using Venn-Peirce diagrams for syllogisms. This paper is divided into three

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 689–696, 2018.
https://doi.org/10.1007/978-3-319-91376-6_61

parts. In the first part, we discuss approaches of those who reject the necessity of fourth figure. In the second part, we review suggestions to incorporate the last figure. In the conclusive part, we develop two key concepts namely – distinct and non-distinct syllogisms along with their equivalence using Venn-Peirce framework. With this, we show that there are exactly three figures.

## 2 The (In)significance of Fourth Figure

Most textbooks of logic [1,2,6,7], etc. report four figures in syllogistic. Aristotle's short-sighted approach, as he could not decipher the last figure is a frequent corollary of the above proposition. This is a historical mistake. It is claimed that as per Aristotle, the position of middle term is the basis of distinction into figures [1,13]. If this is true, then he must presuppose, the predicate of conclusion is the major term and the subject of conclusion is the minor term. But, this presupposition is not found in Aristotle [4].

In fact, Aristotle never gave a general definition of middle term. On the contrary, his definition of middle term is different for each figure [10]. This has perplexed many commentators on Aristotle, as it posed challenge to identify the middle term. Later, Alexander of Aphrodisias suggested that major term is the predicate of that term whose provability is to be investigated, and afterwards Philoponus' account, that the major term is the predicate of the conclusion became widely accepted [8]. Thus, the subject of conclusion becomes minor term and the common term of both premises becomes the middle term.

A way to defend Aristotle's three figures in place of four, is to understand his usage of letters for propositions and syllogisms [16]. He represented propositions with two letters with predicate at the left and subject at the right (e.g. PS) and syllogisms with three letters (say, PMS). Thus, PMS, MPS, and PSM are the three combinations possible (as P will always come before S in this convention). It is just a matter of convention though, as Aristotle has no qualms stating the minor premise first [16]. The number of figures is limited to three because the order of premises is such that the major premise must precede the minor. In order to explain this, many scholars also conjecture that he perhaps used the following diagram[1] as found in many early commentaries [8].

Here, each upper link between the letters is taken to represent the relation of terms expressed in a premise, and each lower link to represent the relation expressed in a conclusion. If Aristotle's thought was guided by such diagrams, then it is easy to see why he assumed that there are only three syllogistic figures [8]. It is clear from the diagrams that there are only three ways in which the

---

[1] Adapted and redrawn from Kneale and Kneale [8].

middle term can be ordered with respect to the two other terms, namely, the major and the minor terms.

Both the above approaches[2] discussed rely heavily on the arrangement of major and minor premise in a syllogism. Along with this, they also show how a relation between major and minor can be inferred by varying the position of middle term. This, in turn opens the debate once again as the position of middle term is crucial in generating figures. If this line of argument is adhered, then it is easier to see that there can be four figures as well. These linguistic approaches appeal to a 'convention' of arrangement to be followed, rather than anything else. There is nothing intrinsic or concrete about conventions in general and (in this case) figures in specific. By defending Aristotle's three figures in this manner, the withstander(s) paves way for the opponent(s), who use the same platform (arrangement of terms and premises) to establish the legitimacy of four figures.

## 3 The Significance of Fourth Figure

In this section, we discuss how this arrangement of terms and premises (as discussed in the last section) produces four figures. As we know that the division of syllogisms into different 'figures' is keenly contested by several scholars. However, the view that syllogisms can be divided into four figures is widespread [1]. Here too, we find that the major premise serves as the first premise and minor as the second premise of the syllogism. This is exactly the same as emphasized by those, who support the three figures theory. The illustration is given below:

| | | | |
|---|---|---|---|
| M — P | P — M | M — P | P — M |
| S — M | S — M | M — S | M — S |
| S — P | S — P | S — P | S — P |
| **Figure 1** | **Figure 2** | **Figure 3** | **Figure 4** |

The above arrangement is elegant, and gives symmetry to the scheme of figures, which was initially lacking in Aristotle's treatment. Furthermore, it is also opined that Aristotle's own work was exploratory but cumbersome in nature. It was later solved and simplified by generations, which followed him [5]. This arrangement of propositions into figures is frequently used by many medieval logicians [9]. Subsequently, it also found place in later works. Several 'moods' occur in these four figures of which, some are valid whereas others are not. In what follows, we enlist valid moods in four figures from the traditional and modern point of view.

### 3.1 Valid Moods in Four Figures (Traditional)

If we take into account the Aristotelian[3] interpretation, there were exactly six syllogisms valid in each figure, as given below:

---

[2] A similar approach is also found in Peterson [12].

[3] 'Aristotelian' here refers to both Aristotle's and his commentators.

| | | | | | | |
|---|---|---|---|---|---|---|
| First Figure | AAA | EAE | AII | EIO | AAI | EAO |
| Second Figure | AEE | EAE | AOO | EIO | AEO | EAO |
| Third Figure | IAI | OAO | AII | EIO | AAI | EAO |
| Fourth Figure | AEE | IAI | AAI | EIO | AEO | EAO |

If we exclude the strengthened and weakened moods, then there are 19 valid moods. They are AAA, EAE, AII, EIO in the first figure; EAE, AEE, EIO, AOO in the second figure; AAI, EAO, IAI, AII, OAO, EIO in the third figure and AAI, AEE, IAI, EAO, EIO in the fourth figure.

### 3.2 Valid Moods in Four Figures (Modern)

There are 15 valid syllogisms in the modern interpretation. Both AAI and EAO in the third and fourth figures are shown to be invalid as they commit 'Existential Fallacy[4]'. The revised list is as follows:

| | | | | | | |
|---|---|---|---|---|---|---|
| First Figure | AAA | EAE | AII | EIO | — | — |
| Second Figure | AEE | EAE | AOO | EIO | — | — |
| Third Figure | IAI | OAO | AII | EIO | — | — |
| Fourth Figure | AEE | IAI | — | EIO | — | — |

The above discussion shows the inconclusive nature of linguistic analysis, since by following any one of the conventions, our ratiocinator can support either of the claim that there are three or four figures. However, Venn-Peirce framework clears this confusion, which we discuss next.

## 4 Venn-Peirce Framework

Venn-Peirce framework [11,21] (imprecisely known as Venn diagrams) is widely used to test the validity of syllogisms. It is now a well proven fact that diagrams are not only effective in human logical reasoning, [18] but also helps in improving our performance and precision [19]. Moreover, it also allows to envision the line of reasoning, which is cumbersome using symbols and becomes complicated in language. In this section, three notions namely, distinct syllogisms, non-distinct syllogisms and diagrammatic congruence are introduced. With this, we argue that the fourth figure is supernumerary.

**Distinct Moods**—A mood is distinct, if it is valid in one figure only. AAA, AOO and OAO are distinct moods in Figs. 1, 2 and 3 respectively.

**Non-distinct Moods**—A mood is non-distinct, if it is valid in more than one figure. EIO is valid in Figs. 1, 2, 3 and 4; EAE in Figs. 1 and 2; AII in Figs. 1 and 3; AEE in Figs. 2 and 4 and IAI is valid in Figs. 3 and 4.

---

[4] A particular conclusion does not follow from two universal premises.

**Diagrammatic Congruence**—Two or more moods are diagrammatically congruent, if they have identical Venn-Peirce diagram. Let us understand this notion, with the help of examples:

1. *EIO*- 1,2,3,4 are represented as:

| EIO-1 | EIO-2 | EIO-3 | EIO-4 |
|---|---|---|---|
| No M is P | No P is M | No M is P | No P is M |
| Some S is M | Some S is M | Some M is S | Some M is S |
| Some S is not P | Some S is not P | Some S is not P | Some S is not P |

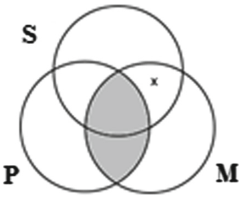


**Fig. 1.** EIO

The ubiquitous EIO in each figure is diagrammatically congruent.

2. *EAE*-1,2 and *AEE*-2,4 are drawn as:

| EAE-1 | EAE-2 | AEE-2 | AEE-4 |
|---|---|---|---|
| No M is P | No P is M | All P is M | All P is M |
| All S is M | All S is M | No S is M | No M is S |
| No S is P | No S is P | No S is P | No S is P |

**Fig. 2.** EAE and AEE

EAE and AEE are not diagrammatically congruent even though they have same combination of premises arranged differently. Nonetheless, EAE-1 and EAE-2 are diagrammatically congruent and the same goes for AEE-2 and AEE-4.

3. *AII*-1,3 and *IAI*-3,4 are diagrammed as:

| AII-1 | AII-3 | IAI-3 | IAI-4 |
|---|---|---|---|
| All M is P | All M is P | Some M is P | Some P is M |
| Some S is M | Some M is S | All M is S | All M is S |
| Some S is P | Some S is P | Some S is P | Some S is P |

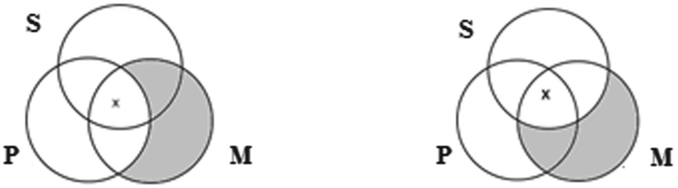


**Fig. 3.** AII and IAI

AII and IAI are not equivalent to each other although, AII-1 and AII-3 are diagrammatically congruent as well as IAI-3 and IAI-4.

4. The distinct moods of *AAA*-1, *AOO*-2 and *OAO*-3 are shown below:

| AAA-1 | AOO-2 | OAO-3 |
|---|---|---|
| All M is P | All P is M | Some M is not P |
| All S is M | Some S is not M | All M is S |
| All S is P | Some S is not P | Some S is not P |

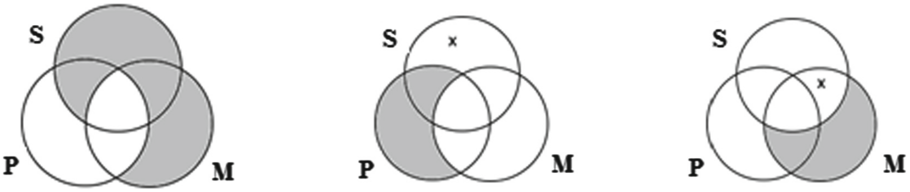


**Fig. 4.** AAA, AOO and OAO

AAA-1, AOO-2 and OAO-3 are the distinct moods.

### 4.1 Summary and Findings

There are exactly 8 (3 distinct and 5 non-distinct) moods that are valid based on the notion of diagrammatic congruence using Venn-Peirce model. They are AAA-1, AOO-2, OAO-3, IAI-3 & 4, AII-1 & 3, AEE-2 & 4, EAE-1 & 2 and EIO-1, 2, 3 & 4. Diagrammatically congruent also mean that they are non-distinct moods present in different figures, which are equivalent[5]. When the position of premises are interchanged (IA & AI and AE & EA), the syllogism so formed look similar though incongruent. There is a distinct mood in every figure except fourth. These findings are important to prove the irrelevance of fourth figure, which we discuss next.

---

[5] The notion of 'Equivalence' is discussed in Richman [15].

### 4.2 Argument Against the Fourth Figure

We have witnessed in Sects. 2 and 3 that the debate on the (in)significance of fourth figure is skewed towards the position of terms. Supporters of three figures end up, calling it a 'matter of convention' whereas admirers of the fourth figure call it, 'elegant and symmetrical'. If the convention of major preceding minor is dropped, we can have eight figures. Furthermore, if we drop the distinction of major and minor terms, we will go on to have sixteen figures. Along with all this, the number of syllogisms valid in all these figures too will increase fourfold. This amplified approach is neither significant nor profound.

Thus, it is prudent to consider something more fundamental. Commentators of Aristotle have earned their livelihood by asking various questions on syllogisms and 'Why Aristotle considered only three figures?' is one of them [3]. Was he really short-sighted, not to see the very obvious fourth figure? Let us answer this question after reconsidering a historical piece of information known to all.

For Aristotle, first figure is the perfect figure[6] and there are four valid syllogisms in it. Why then, are there valid syllogisms in other figures which are (im)perfect? They can be reduced to first figure syllogisms after elementary transformations. If they can be reduced to first figure syllogisms, why is it an imperative to consider them as a separate figure at the very first place? *It contains something unique and signifies a different and robust arrangement of premises and conclusion so formed, to produce an inference in its own way.* There lies his reason of treating middle term in three unequal ways and his [non]-short-sighted approach of ignoring the very obvious fourth figure.

Reasoning with Venn-Peirce diagrams allows us to visually comprehend distinct syllogisms. Additionally, the notion of diagrammatic congruence helps us avoid unnecessary iterations of the same syllogisms. An arrangement of premises and conclusion is separate, when it produces a distinct syllogism. If not, they are non-distinct syllogism(s), that is/are already been considered elsewhere. Thus, a figure is significant and admissible, if and only if it has a distinct syllogism. Absence of a distinct syllogism declares the fourth figure – redundant.

## 5 Conclusion

This paper is a diagrammatic justification to dispense with the fourth figure using Venn-Peirce framework. Also, it shows the efficacy of diagrammatic techniques to follow the principle of parsimony. The fourth figure neither adds anything substantial nor is a cornerstone of syllogistic from any standpoint. Thus, exclusion of it is justified and takes away nothing. In the wake of above development, we reduce the number of valid syllogisms to eight, applying the differentiation of distinct and non-distinct moods. The scope of this paper is limited to explore the notions of distinct, non-distinct and diagrammatic congruence of valid syllogisms only. Application of these concepts to invalid syllogisms will definitely serve as an avenue for discourse(s), but falls beyond the scope of present work.

---

[6] A diagrammatic justification of perfect syllogisms is demonstrated in Sharma [20].

**Acknowledgment.** This paper has substantially benefited from the discussions with Prof. A.V.J. Pietarinen, Dr. G. Stapleton, Dr. J. Burton, Dr. F. Bellucci, Prof. M. Chakraborty and Prof. L. Choudhury, on an earlier draft. I also thank anonymous reviewers for their comments and suggestions.

## References

1. Cohen, M.R., Nagel, E.: An Introduction to Logic and Scientific Method. Allied Publishers, New Delhi (1936)
2. Copi, I.M., Cohen, C.: Introduction to Logic, 8th edn. Macmillan, New York (1990)
3. Flannery, K.L.: Ways into the Logic of Alexander of Aphrodisias. Brill, Leiden (1995)
4. Hadgopoulos, D.J.: The principle of division into four figures in traditional logic. Notre Dame J. Formal Logic **20**, 92–94 (1979)
5. Henle, P.: On the fourth figure of the syllogism. Philos. Sci. **16**, 94–104 (1949)
6. Hurley, P.J.: A Concise Introduction to Logic. Wadsworth, Boston (2012)
7. Keynes, J.N.: Studies and Exercises in Formal Logic. Macmillan, London (1906)
8. Kneale, W., Kneale, M.: The Development of Logic. OUP, London (1962)
9. Łukasiewicz, J.: Aristotle's Syllogistic: From the Standpoint of Modern Formal Logic. OUP, London (1955)
10. McKeon, R.: The Basic Works of Aristotle. Random House, New York (1941)
11. Peirce, C.S.: Collected Papers. Harvard University Press, Cambridge (1933)
12. Peterson, J.B.: The forms of the syllogism. Philos. Rev. **8**, 371–385 (1899)
13. Prior, A.N.: Formal Logic. Clarendon Press, Oxford (1962)
14. Rescher, N.: New light from arabic sources on Galen and the fourth figure. J. History Philos. **3**, 27–41 (1965)
15. Richman, F.: Equivalence of syllogism. NDJFL **45**, 215–233 (2004)
16. Rose, L.E.: Aristotle's syllogistic and the fourth figure. Mind **74**, 382–389 (1965)
17. Sabra, A.I.: A twelfth century defence of the fourth figure of syllogism. J. Warburg Courtauld Inst. **28**, 14–28 (1965)
18. Sato, Y., Mineshima, K.: How diagrams can support syllogistic reasoning: an experimental study. J. Logic Lang. Inf. **24**, 409–455 (2015)
19. Sato, Y., Mineshima, K., Takemura, R.: The efficacy of Euler and Venn diagrams in deductive reasoning: empirical findings. In: Goel, A.K., Jamnik, M., Narayanan, N.H. (eds.) Diagrams 2010. LNCS (LNAI), vol. 6170, pp. 6–22. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14600-8_6
20. Sharma, S.S.: Perfect syllogisms and the method of minimal representation. In: CEUR Workshop Proceedings, vol. 1132, pp. 17–22 (2014)
21. Venn, J.: On the diagrammatic and mechanical representation of propositions and reasonings. Philos. Mag. **10**, 1–18 (1880)

# Posters

# Towards Diagram-Based Editing of Ontologies

Peter Chapman(✉) and William Roberts

Edinburgh Napier University, Edinburgh, UK
`p.chapman@napier.ac.uk`, `10017537@live.napier.ac.uk`

**Abstract.** Ontology creation and editing involves multiple stakeholders, not all of whom may be mathematically trained. Whilst ontology editors, such as Protege, are extensible with visualisation tools to enhance understanding of the ontology, these tools are static representations only. We present initial work on creating editable visualisations for a fragment of OWL, which will in turn update the underlying ontology. The diagrams used are linear diagrams, which have previously been shown to aid comprehension of set-based data. In particular, we focus only on those OWL statements which do not include properties or datatypes.

**Keywords:** Interactive · Linear diagrams · Ontology

## 1 Introduction

Ontologies are a way of reasoning about data in an efficient manner. Ontologies are increasingly prevalent in a range of applications, including the Semantic Web, medicine and law. The development and maintenance of ontologies are skilled tasks requiring knowledge of logical reasoning and symbolic notations. However, the wide variety of stakeholders for each ontology may not have the necessary skill set to perform ontology engineering effectively. Given the critical systems in which ontologies are used, it is of paramount importance that the ontologies encode exactly the information intended.

To alleviate the difficulties in understanding ontologies, various visualisations have been proposed [4]. Some have limited expressiveness, but are implemented as software tools, such as CropCircles [13], and OWLViz [3]. Others are more expressive, but are as yet not implemented, such as concept diagrams [11] and VOWL [6]. Of the existing visualisations, however, all are static, with the exception of OntoTrack [5]. In that visualisation, subclasses and superclasses can be added and deleted from the graphical view. The ontology itself is represented as a directed graph, and thus shows only subsumption relationships directly. Thus, features of the hierarchy such as disjointness of classes, or subsumption under a union, are not explicitly represented.

---

W. Roberts was supported by an Edinburgh Napier University Internal Research Grant.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 699–703, 2018.
https://doi.org/10.1007/978-3-319-91376-6_62

We propose, and have completed initial work on, inLineD (for **in**teractive **Line**ar **D**iagrams) an ontology visualisation tool that allows users to: edit how the ontology hierarchy is represented; add, remove and edit new classes, including intersection and union information; and add, remove and edit disjointness axioms.

## 2 Adding Interactivity to Linear Diagrams

Linear diagrams are originally due to Leibniz [2], and use horizontal line segments to represent sets. The vertical alignment of the segments represents intersection properties amongst the sets. For example, consider Fig. 1, representing the interests of people in a social network. We see six sets represented, using 14 overlaps and 10 line segments. The fifth overlap from the left, consisting of line segments in the first, third and fourth rows, encodes the information that the set $\text{Relaxation} \cap \text{Media} \cap \text{Design} \cap \overline{\text{Cars}} \cap \overline{\text{Travel}} \cap \overline{\text{Health}}$ is non-empty. Similarly, since every overlap containing a line segment in the sixth row also contains a line segment in the first row, we can infer that $\text{Health} \subseteq \text{Relaxation}$. Finally, since there is no overlap containing line segments from both the fifth and sixth rows, we can infer that $\text{Travel} \cap \text{Health} = \emptyset$.

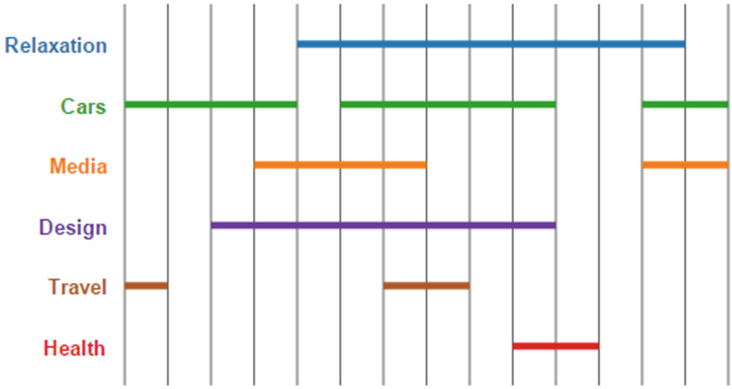


**Fig. 1.** A linear diagram

Figure 1 was produced using the predecessor of our tool, the Linear Diagram Generator [8]. The layout of the diagram is based upon the layout principles of [9]: the total number of line segments is minimised; there are vertical guidelines to aid reading overlaps; the lines are each of a single colour; the lines themselves are relatively thin; the orientation of the diagram is horizontal; and the vertical order of the sets is adjacency-driven. By the last, we mean that two sets which have segments which begin and end next to each other (as between the fourth and fifth overlaps for the lines Relaxation and Cars) are drawn as close as possible to each other. In this example, they are drawn next to each other.

These layout principles were shown in [9] to aid in answering general set-based questions. However, the user may want to ask questions for which this layout is not optimal. Interaction in a visualisation can allow the user to interrogate the underlying data more effectively [7]. For example, consider the question "is everyone interested in Health also interested at least one of Media and Cars?" By redrawing Fig. 1 using inLineD to draw the sets Cars and Media as single segments, and bringing the three sets of interest next to each other, the answer to the question (no) is more readily apparent, as shown in Fig. 2. This redesign uses 13 line segments, compared to the original's 10, and is thus sub-optimal *in general*. In the context of ontologies, by giving the ontology engineer facility to redraw the diagram, specific relationships amongst small numbers of classes can be closely analysed, and, as we shall see in the next section, manipulated.

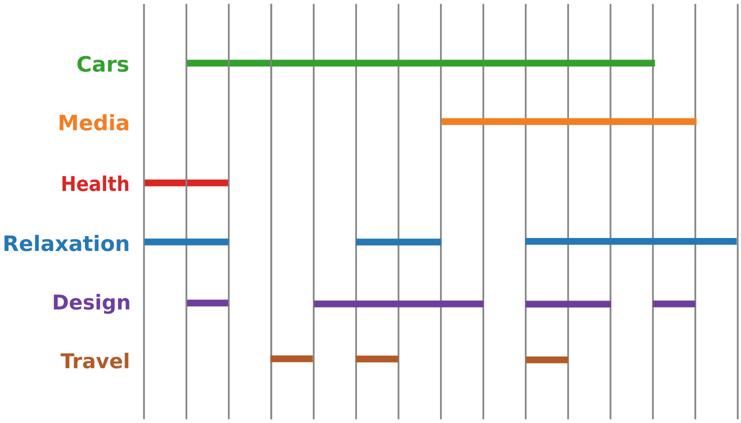


**Fig. 2.** A redrawn version of Fig. 1

## 3 Next Steps - Heterogeneous Reasoning

Manipulating diagrams, unlike the example in the previous section, may sometimes change the semantic content of the diagram. For example, consider Fig. 3, the top diagram of which is Venn-2. Suppose that this arrangement was *not* what the ontology engineer was expecting. By removing the overlap where the two sets $A$ and $B$ both have line segments, we would have introduced the fact that $A \cap B = \emptyset$ (bottom left of Fig. 3). Similarly, by removing every overlap where $A$ has line segments but $B$ does not would introduce the fact that $A \subseteq B$ (bottom right of Fig. 3). These two types of statement form the basis of OWL axiom schemes, as $\mathsf{disjoint}(A, B)$ and $A$ $\mathsf{is\ A}$ $B$ respectively [1]. By manipulating the diagram, then, the engineer could introduce new axioms into the ontology. Adapting and implementing the translations of [10,12] to the case of linear diagrams would then create a heterogeneous system: axioms could be entered as

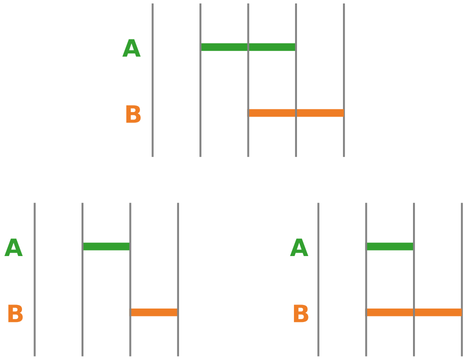


**Fig. 3.** Manipulations that change the underlying semantics

text, and the diagram would update automatically; conversely, changing the diagram would update the set of axioms. The move to such as a system is the next stage for inLineD.

## References

1. Baader, F., Calvanese, D., McGuinness, D., Nadi, D., Patel-Schneider, P. (eds.): The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, Cambridge (2003)
2. Couturat, L.: Opuscules et fragments inédits de Leibniz. Felix Alcan, Paris (1903)
3. Horridge, M.: OWLViz. http://protegewiki.stanford.edu/wiki/OWLViz. Accessed Aug 2014
4. Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., Giannopoulou, E.: Ontology visualization methodsa survey. ACM Comput. Surv. (CSUR) **39**(4), 10 (2007)
5. Liebig, T., Noppens, O.: OntoTrack: combining browsing and editing with reasoning and explaining for OWL lite ontologies. In: McIlraith, S.A., Plexousakis, D., van Harmelen, F. (eds.) ISWC 2004. LNCS, vol. 3298, pp. 244–258. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30475-3_18
6. Lohmann, S., Negru, S., Haag, F., Ertl, T.: Visualizing ontologies with VOWL. Semant. Web **7**(4), 399–419 (2016)
7. Munzner, T.: Visualization Analysis and Design. CRC Press, Boca Raton (2014)
8. Rodgers, P.: Linear diagram generator (2017). https://www.cs.kent.ac.uk/people/staff/pjr/linear/index.html. Accessed Dec 2017
9. Rodgers, P., Stapleton, G., Chapman, P.: Visualizing sets with linear diagrams. ACM Trans. Comput.-Hum. Interact. (TOCHI) **22**(6), 27 (2015)
10. Stapleton, G., Masthoff, J.: Incorporating negation into visual logics: a case study using Euler diagrams. In: Visual Languages and Computing 2007, pp. 187–194. Knowledge Systems Institute (2007)
11. Stapleton, G., Howse, J., Chapman, P., Delaney, A., Burton, J., Oliver, I.: Formalizing concept diagrams. In: Visual Languages and Computing, pp. 182–187 (2013)

12. Swoboda, N., Allwein, G.: Heterogeneous reasoning with Euler/Venn diagrams containing named constants and FOL. In: Proceedings of Euler Diagrams 2004. ENTCS, vol. 134. Elsevier Science (2005)
13. Wang, T.D., Parsia, B.: CropCircles: topology sensitive visualization of OWL class hierarchies. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 695–708. Springer, Heidelberg (2006). https://doi.org/10.1007/11926078_50

# Diagrams Including Pictograms Increase Stock-Flow Performance

Friederike Brockhaus(✉) and Peter Sedlmeier

Chemnitz University of Technology, 09107 Chemnitz, Germany
{friederike.brockhaus,
peter.sedlmeier}@psychologie.tu-chemnitz.de

**Abstract.** Stock-flow (SF) systems are omnipresent in our lives while difficult to understand. An example is the amount of $CO_2$ in the atmosphere (stock) that changes in dependence of incoming $CO_2$ (inflow) and outgoing $CO_2$ (outflow). When participants are to deal with such tasks, they show poor performance. Despite several attempts to facilitate SF knowledge in participants, as far as we know, only one manipulation led to meaningfully increased SF performance: Changing the representation of the flows into pictograms. In the current study, we intend to modify these kind of diagrams so that they communicate SF information in a simple way. We tested whether the modified representation triggered basic SF understanding. Each participant worked on two tasks; one shown as line graph and one shown as diagram with pictograms. Getting the pictograms at first position led to strongly improved SF performance. A t-test revealed more correct solutions for pictograms than for line graph at first position. Again, the representation of the flows as pictograms led to better SF performance.

**Keywords:** Stock-flow systems · Representation format · Pictograms

## 1 Introduction

Since SF systems are omnipresent in our personal lives, it is very important to understand them correctly. When people have to solve SF problems, they often fail, even if they are highly educated [1]. They seldom understand the "principle of accumulation": as long as the inflow is larger than the outflow, the stock is increasing and vice versa [2]. For example, participants received a graph of water flowing in and out of a water tub. They had to draw the changes of the stock into an empty graph. 40% to 70% or more were committing errors [1–4]. Authors from different backgrounds tried to improve participants' problem-solving performance. However, up to now, almost no meaningful amelioration has been found [2, 5].

Brockhaus et al. [6] examined if changing the representation format of the flow graphs led to better SF performance. Indeed, using diagrams including pictograms instead of line graphs increased SF performance: 57% to 83% solved the tasks correctly compared to 23% to 30% in the baseline condition with line graphs ($\beta$ = .59 to $\beta$ = .53 with $R^2$ = .39; $F(5,84) = 10.86$; $p < .001$). Another study was conducted in order to find out, whether the improved performance still occurred when subjects answered

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 704–707, 2018.
https://doi.org/10.1007/978-3-319-91376-6_63

questions instead of drawing the changes of the stock [7]. This way, a basic SF knowledge was tested, since participants only judged the maximum and minimum of the stock instead of calculating the stock for each moment. Performance was still better for the diagrams including pictograms than for the line graphs, but the effect decreased strongly. When line graphs were given, about a third of the participants solved the tasks correctly, whereas in the conditions with pictograms, they ranged from 41% to 53% ($\beta = .20$ with $R^2 = .22$; $F(10{,}120) = 3.424$; $p < .001$).

The aim of the present pilot study was to check whether basic SF knowledge can be triggered, when we change the format of the diagram performance following Neurath's [8] suggestions to create such diagrams. Our hypothesis was that the new flow-diagram including pictograms led to meaningfully better performance than the line graphs.

## 2 Method

Twenty-four undergraduates from Chemnitz University of Technology (58% female; mean age: 22 years, *SD* = 4.02) participated in a paper-pencil test and received course credit. Every participant had to work on two tasks. Each task included two questions: When is the stock at its maximum and when at its minimum? Answers were coded as correct or incorrect. The inflow- and outflow-patterns were identical for the two tasks but the representation format was changed: Line graphs (Fig. 1) or pictograms (Fig. 2) showed the flows. We varied the *position* of the line diagrams and pictograms to work on: half of the participants worked on the line graph first and then on the diagram including pictograms, the other half started with pictograms.

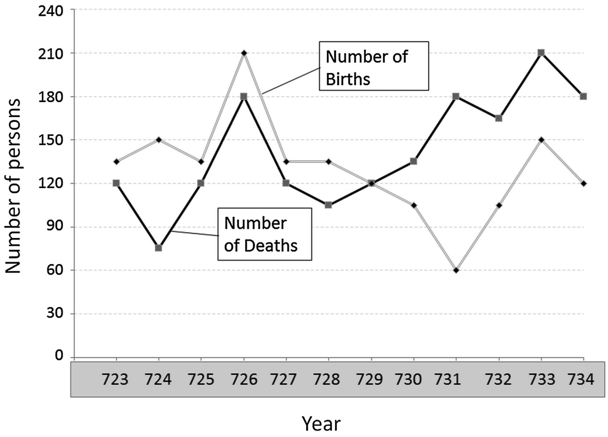


**Fig. 1.** Conventional line graph of the flows

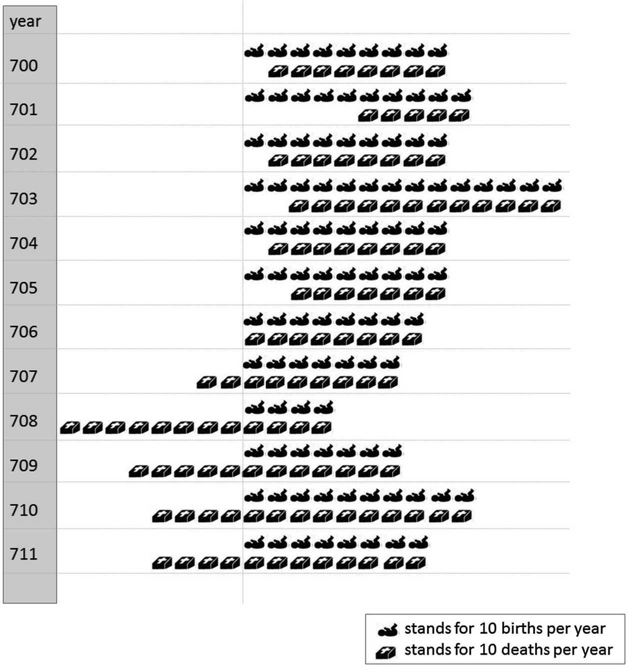


**Fig. 2.** Diagram including pictograms of the flows [9]

## 3 Results

SF performance was rather bad in all conditions, ranging from 8% to 33% of correct solutions. When we took the position of the diagrams into account, the picture changed (Fig. 3).

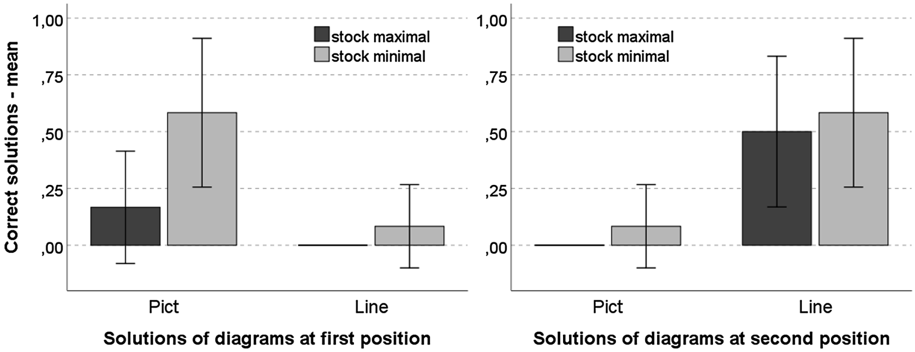


**Fig. 3.** Amount of correct solutions (mean percent correct) and 95% Confidence Intervals for both diagrams and both questions concerning the stock, separated into the group being at first position vs. the group being at second position (legend: Line = line graph, Pict = pictograms)

Participants who worked on the line graph first showed bad performance for both tasks. Performance was much better for those who worked on the pictograms first. Comparing the SF performance for line graph at first position vs. pictograms at first position showed a significant better solutions for the pictograms ($t(22) = 2.861$, $p = .01$ and $d = 1.17$).

## 4 Discussion

Pictograms at first position led to better SF performance than line graphs at first position. Because of the small sample, it is necessary to replicate the study to see whether the effect is stable. Furthermore, the flow-patterns of the two tasks were identical and the solutions' correctness of the two tasks correlated ($r = .89$). Therefore, it was unclear, if one group was better in solving SF tasks, if participants realized that the two tasks were identical, or if they used the same solving approach for both tasks. Further studies should control these possibilities. Up to now, the present data suggests that pictograms support SF performance.

## References

1. Booth Sweeney, L., Sterman, J.D.: Bathtub dynamics: initial results of a systems thinking inventory. Syst. Dyn. Rev. **16**(4), 249–286 (2000)
2. Cronin, M.A., Gonzalez, C., Sterman, J.D.: Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. Organ. Behav. Hum. Decis. Process. **108**(1), 116–130 (2009)
3. Ossimitz, G.: Stock-flow-thinking and reading stock-flow-related graphs: an empirical investigation in dynamic thinking abilities. In: Paper presented at the 20th International Conference of the System Dynamics Society (2002). http://www.systemdynamics.org/conferences/2002/proceed/papers/Ossimit1.pdf. Accessed 06 Dec 2017
4. Kapmeier, F., Happach, R.M., Tilebein, M.: Bathtub dynamics revisited: an examination of déformation professionelle in higher education. Syst. Res. Behav. Sci. **34**(3), 227–249 (2017)
5. Schwarz, M.A.: Vom stockenden Verständnis fließender Zusammenhänge. Darstellungs- und personenbezogene Einflussfaktoren auf das basale Verständnis einfacher dynamischer Systeme (2016). http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-203692. Accessed Mar 2018
6. Brockhaus, F., Arnold, J., Schwarz, M.A., Sedlmeier, P.: Does the modification of the representation format affect stock-flow-thinking? In: Paper presented at the 31st International Conference of the System Dynamics Society (2013). http://www.systemdynamics.org/conferences/2013/proceed/papers/P1253.pdf. Accessed 06 Dec 2017
7. Brockhaus, F.: Von den Flussgrößen zum Bestand: Lassen sich einfache dynamische Systeme intuitiv verständlich darstellen? (2018, to be published)
8. Neurath, O.: International Picture Language (1936). https://monoskop.org/images/e/ec/Neurath_Otto_International_Picture_Language.pdf. Accessed 26 June 2017
9. Arntz, G.: Pictograms (Babies and coffins) / © VG Bild-Kunst, Bonn (2018)

# A Simple Decision Method for Syllogistic

Frank Thomas Sautter[1](✉) and Gisele Dalva Secco[2]

[1] Universidade Federal de Santa Maria, Santa Maria, Brazil
ftsautter@ufsm.br
[2] Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
gisele.secco@ufrgs.br

**Abstract.** We present a method of digraphs for Syllogistic that uses only two rules for testing the validity of syllogisms without existential import and a third rule for cases in which the existential import of terms is used. This method derives from Martin Gardner's network method for Classical Propositional Logic, preserving the iconicity features that Gardner attributes to the propositional case also in the case of Syllogistic. We will first present the graphical representations and the rules for manipulating these representations in the case of syllogisms in which the existential import of terms are not admitted. Then, we will extend the method with a graphical representation for the existential import and with a new rule for the manipulation of this representation. Finally, we will show some applications of the method. It was first presented in Portuguese in *Cognitio* 14(2013): 221–234.

**Keywords:** Digraph · Martin Gardner · Supreme Rules of Inference

## 1 Introduction

Gardner [1] examined a network method for Classical Propositional Logic, and later formulated a variant that uses digraphs [2]. In spite of asking about the possibility of applying the technique to other logical systems, he was unable to apply it to logical systems as simple as Syllogistic. Sautter [3] developed a variant of Gardner's Method for Syllogistic. This method preserves the iconicity features that Gardner attributes to his own method applied to the propositional case.

First, we will present the graphical representations and the rules for manipulation of these representations to the syllogisms in which the existential import of the terms are not admitted. Then, we will extend the method with the graphical representation of the existential import, as well as with a new rule for the manipulation of this graphical representation. Finally, we will show briefly some applications of the method.

## 2 The Basics of the Method

Our method implements the conceiving of Syllogistic as a theory of subordination of concepts. They occur in pairs, for instance the concept X and its related con-

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 708–711, 2018.
https://doi.org/10.1007/978-3-319-91376-6_64

cept Other-than-X (represented as $\bar{X}$). Pairs of arrows expresses a subordination and a single straight line expresses a non-subordination. “All X is Y” expresses the subordination of the concept X to the concept Y; “No X is Y” expresses the subordination of the concept X to the concept Other-than-Y; “Some X is Y” expresses the non-subordination of the concept X to the concept Other-than-Y; and “Some X is not Y” expresses the non-subordination of the concept X to the concept Y (See Fig. 1).

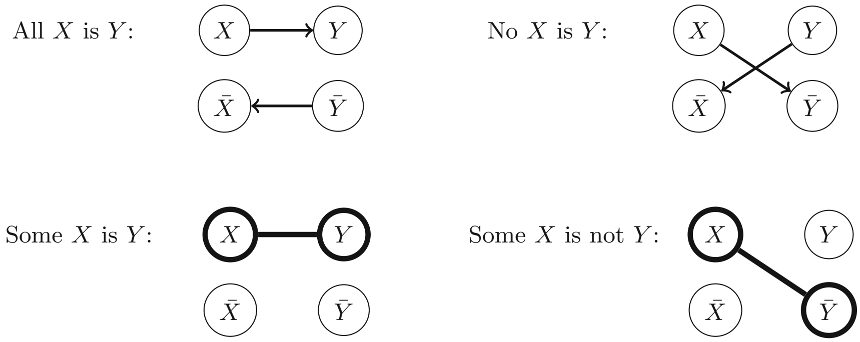


**Fig. 1.** Categorical propositions

Traditionally, there are two supreme rules of rational inference, distinguished one from another by the quality of the categorical propositions: one rule for syllogisms only with affirmative premises and the other rule for syllogisms with affirmative and negative premises. Our method also operates only with two rules for the validity of syllogisms without existential presupposition, but they are distinguished one from another by the quantity of the categorical propositions: one rule for syllogisms only with universal premises and the other rule for syllogisms with quantitatively mixed premises, i.e. one premise is universal and the other is particular (See Fig. 2). This implies the existence of four fallacious rules: two of them related to the cases in which the arrows move in the opposite direction, one of them in which a straight line follows (and not precedes) an arrow, and the last of them in which there are two straight lines.

Diagrams show the artificiality of the division into figures and modes. There are only eight distinct diagrams that stand for valid presuppositionless syllogisms: a diagram for BARBARA; a diagram for CELARENT and CESARE; a diagram for DARII and DATISI; a diagram for FERIO, FESTINO, FERISON and FRESISON; a diagram for CAMESTRES and CAMENES; a diagram for BAROCO; a diagram for BOCARDO; and a diagram for DISAMIS and DIMARIS.

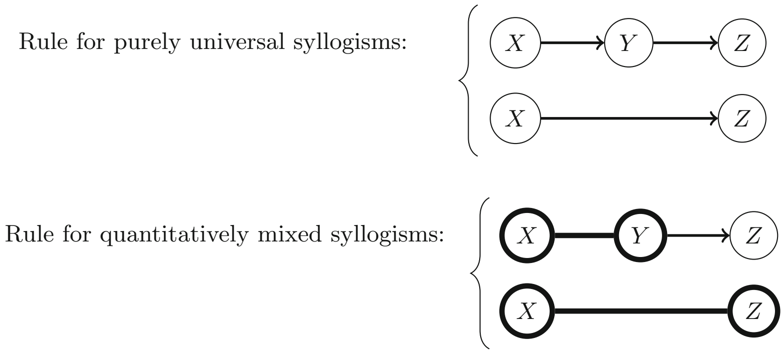


**Fig. 2.** Rules of inference

## 3 Existential Import

Figure 3 shows how to represent the existential presupposition of a term and also a new rule applicable to such cases. If existential pressuposition is admitted, there are six new diagrams for valid syllogisms: one of them stands for BARBARI; other stands for CELARONT and CESARONT; other stands for CAMESTROS and CAMENOP; other stands for DARAPTI; other stands for FELAPTON and FESAPO, and other stands for BRAMANTIP.

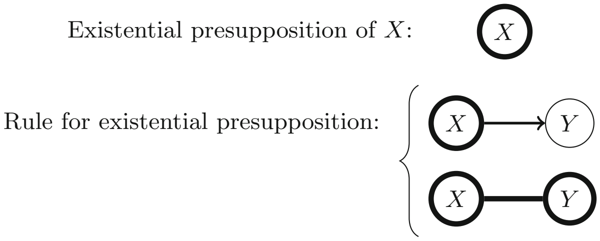


**Fig. 3.** Existential presupposition

## 4 Applications

Figure 4 shows several applications of the method. Figure 4(a) shows the validity of the syllogism EIO of the First Figure, because it is possible to move from the premises to the conclusion by applying the rule for quantitatively mixed premises. No valid rule can move us from the premises to the conclusion in the invalid syllogism IAI of the First Figure represented by Fig. 4(b). Figure 4(c) shows BARBARI; the existential import of the term $S$, followed by the application of the rule for existential presupposition, followed by the rule for quantitatively mixed premises are required to verify its validity. Finally, Fig. 4(d) shows a pair of premises; "Some S is not P" can be concluded. Given a premise and the conclusion, the method can also be easily applied to find the missing premise of a valid syllogism when it exists.

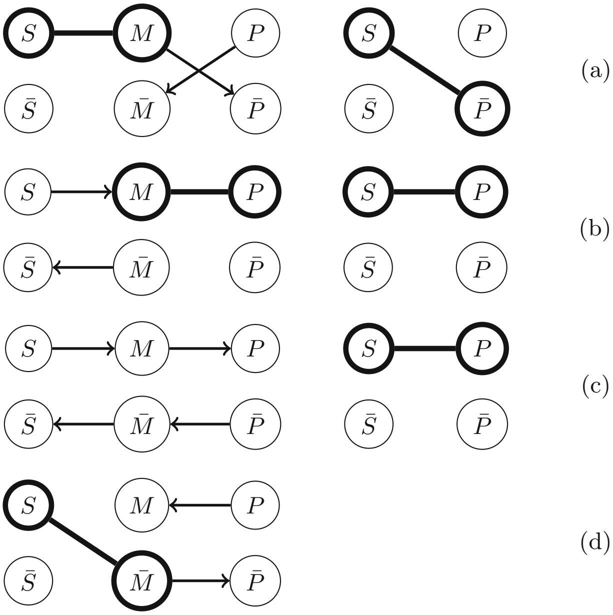


**Fig. 4.** Some applications of the method.

## References

1. Gardner, M.: Network diagram for the propositional calculus. In: Logic Machines and Diagrams, pp. 60–79. McGraw Hill, New York (1958)
2. Gardner, M.: The propositional calculus with directed graphs. In: A Gardner's Workout: Training the Mind and Entertaining the Spirit, pp. 25–33. A. K. Peters, Natick (2001)
3. Sautter, F.T.: Gardner's method for syllogistic. Cognitio **14**, 221–234 (2013). (in Portuguese)

# Wallis's Use of Innovative Diagrams

Erika Rita Ortiz[1,2(✉)]

[1] National University of Cordoba, Cordoba, Argentina
erikarortiz@gmail.com
[2] National Research Council (CONICET), Buenos Aires, Argentina

**Abstract.** Mathematical research is best characterized by problem-solving activities which make use of a variety of modes of representation. Against this background, my aim is to discuss the epistemic value of diagrammatic representation in problem-solving. To make my point, I consider a case study selected from Wallis's work on the quadrature of conic sections. Wallis's definition of conic sections is given in terms of algebraic equations setting them free from 'the embrangling of the cone'. This suggests the aim to eliminate figures and other iconic elements with a view to attaining higher level of abstraction but, in Wallis's work, geometric diagrams display relations that can be fruitfully used to calculate arithmetically the area of a figure. The use of displayed relations leads to the formulation of algebraic equations defining curves and it is also what makes room for arithmetical calculations. Accordingly, the notion of a general method of resolution is grounded on properties read off the diagram so that despite Wallis's insistence on algebraic representation -I argue- diagrams remain essential working tools.

**Keywords:** Conic sections · Diagrams · Algebraic equations · Wallis

## 1 Introduction

Twenty century philosophy of mathematics focused on the foundations of axiomatic systems and formal arguments, assuming that mathematical reasoning is best characterized as purely syntactic without any reference to the context of work. The syntactic presentation makes explicit all relevant steps in a proof guaranteeing formal rigor. From such assumptions follows that figures and, more generally, iconic ingredients are to be eliminated from a fully systematic presentation. In accordance with the rejection of any reference to context of work the specificity of problem solving activities have been eliminated.

More recently, and despite the apparent persistence of this perspective, some critical voices have been insisting upon the value of paying closer attention to the work of the research mathematician focusing on the specificity of his problem-solving activities in the relevant context of work. Following Grosholz [1] we argue that mathematical research is best characterized by problem-solving activities which make use of a variety of modes of representation such as symbolical notations, equations and diagrams. Against this background, my aim is to discuss the

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 712–715, 2018.
https://doi.org/10.1007/978-3-319-91376-6_65

epistemic value of diagrammatic representation in problem-solving. To make my point, I will consider a case study selected from Wallis's work on the quadrature of conic sections.

## 2 Defining the Conics: From the Cone to Algebraic Equations

The goal underlying Wallis's exposition in *Sectionibus Conicis* (SC) and *Arithmetica Infinitorum* (AI) (1656) is to find a general method to solve quadrature problems. In the SC Wallis gives a systematic treatment of conic sections that culminates in the definition of these figures in terms of algebraic equations. The aim of these definitions developed in SC is to determine the properties needed to solve quadrature problems by means of arithmetical series. On the other hand, in AI Wallis offers a general method to solve quadrature problems based on the definitions given in the first text. Wallis was aware of the limitations of purely geometrical approaches to solve quadrature problems by means of a general method. In his research he acknowledges the contribution of the method of indivisibles as well as Descartes's work in geometry. His recognition of the value of Cavalieri's contribution is exhibited in the first proposition of SC where Wallis explicitly appeals to the method of indivisibles: "I suppose at the start (according to the *Geometria indivisibilium* of Bonaventura Cavalieri) that any plane whatever consists, as it were, of an infinite number of parallel lines. Or rather (which I prefer) of an infinite number of parallelograms of equal altitude; which indeed the altitude of a single one is $\frac{1}{\infty}$ of the whole altitude, or an infinitely small divisor; (...); and therefore, the altitude of all of them at once is equal to the altitude of the figure". [2, p. 68] [4, p. 4]

Once that Wallis has established the frame of his investigation, he next considers the parabola, the ellipse and the hyperbola as sections of a cone. Following Descartes, Wallis considers these curves not as sections of a cone, but as plane figures (Fig. 1) susceptible of being described by algebraic equations. To find the equations defining the conics Wallis examines the geometrical relations (*ratios*) among the elements the curves consist of. For the parabola he considers ordinates of the curve, the *latus rectum* and the diameters (DO, LA and AD respectively in Fig. 1) [4, pp. 46–59]. Focusing on the relation between the ordinates and diameters, Wallis derives the equation of the parabola $p^2 = ld$.

We note that Wallis's treatment of conics relies on the work of Apollonius, while he considers these curves as sections of a cone and retains the geometrical terminology for its constitutive parts. But differences between them become clearer, as soon as he incorporates elements from other mathematical traditions which he articulates in the diagram that accompanies his exposition. Diagrams now represent curves re-conceived as planar figures. Following Cavalieri's method, he takes planar figures to consist of an infinite number of parallelograms having indefinitely small altitude. Moreover, according to Descartes, such curves are susceptible of description by algebraic equations. The elements thus articulated in diagrams, exceed the synthetic and finitary character of classical

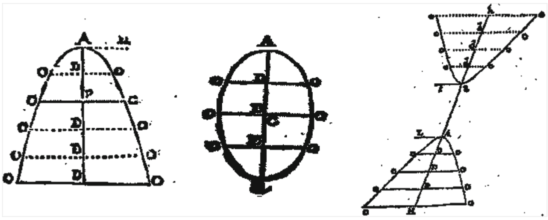


**Fig. 1.** Conic section as planar figures [4, p. 63, 60 and 79].

geometry showing to be fundamental for the determination of properties needed to solve quadrature problems in a general way. With the aim to evaluate the role of these innovative diagrams in the next section I will present the resolution of the quadrature of the parabola.

## 3 The Arithmetical Quadrature of the Parabola

The first stage in the determination of the quadrature of the parabola aims to calculate the ratio between the series of square numbers and the series with the same number of terms equal to the greatest term of first series. The investigation proceeds by induction and the result is reached in proposition 21 which affirms that "If there is proposed an infinite series, of quantities that are as squares of arithmetic proportionals continually increasing (...) it will be to a series of the same number of terms equal to the greatest as 1 to 3". [3, p. 27] After establishing this arithmetical result Wallis focuses on the determination of the area under the parabola. Wallis formulates the problem referring explicitly to the diagram that goes with the text (Fig. 2). The resolution requires to determine the proportion between the area of AOT and the rectangle ATOD. As the diagram shows, both figures are considered according to Cavalieri's method as consisting of indivisibles - lines DO and TO. The initial problem is thus transformed into the problem to determine the proportion between collection of indivisibles. To determine the required proportion Wallis associates numerical values to indivisibles composing both figures. In this case, the initial terms of both series are put in correspondence with the vertex of the figure (point A), while indivisibles of AOT correlate with terms of the series of square numbers and indivisibles of ATOD with terms of the series whose terms are equal to the greatest term of first series. The correspondence between numerical values and indivisibles can only be established if the relations (ratio) between terms of series is the same as the one that holds between ordinates and diameter of the curve. Relations between ordinates and diameters were already examined by Wallis - we recall - when focusing on the definition of conics by algebraic equations (Sect. 2). Here Wallis refers to previous results which establish that straight lines DO (ordinates) are as the square roots of the lines AD (diameters) and conversely lines AD - that is, TO - will be as the squares of the same DO - that is, AT [4, pp. 46–50]. This result guarantees the correspondence between numerical values and indivisibles

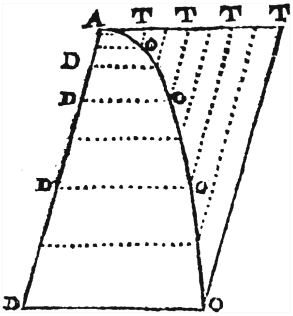


**Fig. 2.** [3, p. 28]

and allows Wallis to apply the arithmetical results reached in preceding propositions: "Therefore the whole figure AOT (...) will be, to the parallelogram of equal height TD (...), as 1 to 3". [3, p. 28]

## 4 Concluding Remarks

In our case study we saw how the diagram introduced by Wallis for the calculation of the area under the parabola does not show any reference to the cone but shows a new way to formulate quadrature problems. Thus, problems about determination of areas are transformed into problems of determination of proportions between collection of indivisibles. Wallis realizes that the required proportion can be achieved by associating indivisibles with numerical values. This association can only be done if the relation between the terms of series is the same as the one that holds between indivisibles of figures. Wallis appeals to the geometrical relations established in his search of algebraic equations defining the conics. Geometric relations on which equations are based thus become conditions of solvability of quadrature problems. To conclude, Wallis's novel formulation of quadrature problems is made possible by the method of indivisibles, while his resolution depends on the Cartesian approach. Both elements collapse in his conception of the conics as planar figures that can only be fully articulated by the diagrams accompanying the text.

## References

1. Grosholz, E.: Representation and Productive Ambiguity in the Mathematics and the Sciences. Oxford University Press, Oxford (2007)
2. Stedall, J.: Mathematics Emerging: A Sourcebook 1540–1900. Oxford University Press, Oxford (2008)
3. Wallis, J.: The Arithmetics of Infinitesimals. John Wallis 1656. Springer, New York (2004). https://doi.org/10.1007/978-1-4757-4312-8. Introduction and translation by J. Stedall
4. Wallis, J.: De sectionibus conicis, nova methodo expositis, tractatus. Leon Litchfield, Oxford (1656)

# Venn Diagram and Evaluation of Syllogisms with Negative Terms: A New Algorithm

Mehdi Mirzapour(✉) and Christian Retoré

LIRMM – UMR 5506, Université de Montpellier and CNRS,
Montpellier, France
`mehdi.mirzapour@gmail.com`

**Abstract.** We propose an algorithmic procedure for the automatic analysis of syllogisms with negative terms based on the modified version of Shin's Venn-I diagram. Our computational procedure can automatically generate all the possible conclusions derivable from the two premises of a given syllogism with negative terms. Our approach relies on the reformulation of the logic behind the relations between points, lines, and surfaces in the Venn diagram by employing conditional propagation rules.

## 1 Introduction

In this paper, we provide a three-step algorithmic procedure for the automatic analysis of syllogisms with negative terms based on the formalization of the modified Shin's Venn-I diagram [3]. Our proposal applies slight changes in the representation of the points, lines and surfaces in the Venn-I diagram for computational convenience (Sect. 2.1). It also introduces 48 conditional propagation rules for displacement of the existential points (Sect. 2.2). In the next section, we briefly describe how these changes are sufficient for the automatic evaluation of syllogisms with negative terms.

## 2 Algorithmic Procedure

Given $X$ and $Y$, (distinct) members of the set of terms $\{S, P, M\}$ (unary predicates in modern logic sense), we can define *Every* $X$ *is* $Y$ as $XaY$; *No* $X$ *is* $Y$ as $XeY$; *Some* $X$ *is* $Y$ as $XiY$ and *Some* $X$ *is not* $Y$ as $XoY$. This definition can be extended to a syllogism that allows to include the negation of the terms $\{S', P', M'\}$. This extension allows to have the equivalent formulas such as $XaY = XeY'$. Moreover, it adds new kind of propositions with negative terms that were absent in the standard definition of the syllogism. Now, we define a syllogism which consists of two premises and one conclusion. The first premise can have two terms $S (or\, S')$ and $M (or\, M')$ with different possible quantifiers; the same thing for the second premise but with the two terms $M (or\, M')$ and

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 716–720, 2018.
https://doi.org/10.1007/978-3-319-91376-6_66

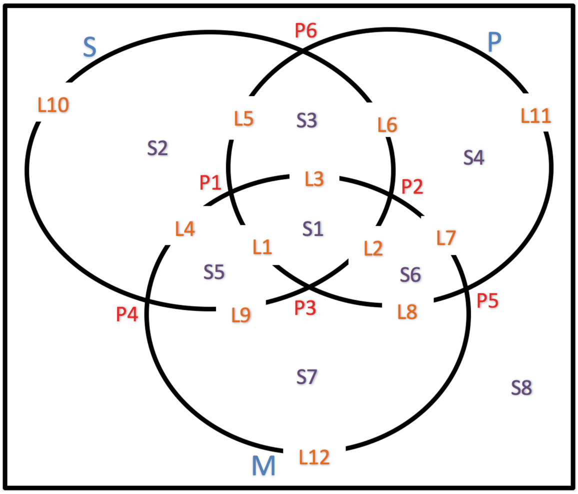


**Fig. 1.** Venn diagram for syllogism with negative terms

$P(or\ P')$. The conclusion has only two terms $S(or\ S')$ and $P(or\ P')$—the middle term $M(or\ M')$ is absent from the conclusion.

Figure 1 illustrates a labeled version of the Venn diagram [2]. As one can observe, all the 6 points, 12 lines and 8 surfaces[1] are labeled by using the indexed variables in the diagram. Given $G = \{P_1, \cdots, P_6, L_1, \cdots, L_{12}, S_1, \cdots, S_8\}$, we define a valuation $V_i$ $(i \in \mathbb{N})$ a mapping from $G$ to $\{0, 1, u\}$ such that valuations of the points and lines are never 0. We interpret 0 as emptiness and 1 as non-emptiness (or existential import) and $u$ as undefined value (stating nothing in the diagram). The formula $V_i(L) = 1$ means that one of the two adjacent surfaces, separated by the line $L$, is 1. The formula $V_i(P) = 1$ means that one of the four adjacent surfaces – which $P$ is the point of intersection between their borders – is 1. The geometrical interpretation can be seen in the legend map of Fig. 2. Having these concepts, we can describe the procedure:

## 2.1 Step I: Valuation

In this paper, we assume existential import for all the terms[2]. Consequently, we can assume the non-emptiness of $S, P, M, P', S', M'$ in the Venn diagram, so, we can state this fact with $V_0(P_i) = 1$ (for $1 \leq i \leq 6$). Given two potential premises of a syllogism, we can define a valuation $V_1$ provided in the rows (1–16) of the

---

[1] The term *surface* is called *zone* in the classical Venn and Euler diagram literature.

[2] Our proposal also works without assuming existential import. We only need to eliminate 24 point-to-line propagation rules introduced in Sect. 2.2. Another possibility is to keep the rules without assigning valuation 1 to the points. Obviously, the combination of those two solutions is also possible.

Table 1. We have $V_1(P_i) = V_0(P_i) = 1$ for $1 \leq i \leq 6$ and the remaining geometric variables are assigned $u$ by $V_1$.

**Table 1.** Venn diagram valuations and conditions

| ID | Equivalent Logical Groups | | | | **Valuation(s)** | ID | Equivalent Logical Groups | | | | **Condition** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **SaM** | SeM' | M'eS | M'aS' | **V1(S2)=0, V1(S3)=0** | 17 | **SaP** | SeP' | P'eS | P'aS' | **S2=0 ∧ S5=0** |
| 2 | **SeM** | SaM' | MeS | MaS' | **V1(S1)=0 ,V1(S5)=0** | 18 | **SeP** | SaP' | PeS | PaS' | **S1=0 ∧ S3=0** |
| 3 | **SiM** | SoM' | MiS | MoS' | **V1(L1)=1** | 19 | **SiP** | SoP' | PiS | PoS' | **S1=1 ∨ S3=1 ∨ L3=1** |
| 4 | **SoM** | SiM' | M'iS | M'oS' | **V1(L5)=1** | 20 | **SoP** | SiP' | P'iS | P'oS' | **S2=1 ∨ S5=1 ∨ L4=1** |
| 5 | **S'aM'** | S'eM | MeS' | MaS | **V1(S6)=0, V1(S7)=0** | 21 | **S'aP'** | S'eP | PeS' | PaS | **S4=0 ∧ S6=0** |
| 6 | **S'eM'** | S'aM | M'eS' | M'aS | **V1(S4)=0, V1(S8)=0** | 22 | **S'eP'** | S'aP | P'eS' | P'aS | **S7=0 ∧ S8=0** |
| 7 | **S'iM'** | S'oM | M'iS' | M'oS | **V1(L11)=1** | 23 | **S'iP'** | S'oP | P'iS' | P'oS | **S7=1 ∨ S8=1 ∨ L12=1** |
| 8 | **S'oM'** | S'iM | MiS' | MoS | **V1(L8)=1** | 24 | **S'oP'** | S'iP | PiS' | PoS | **S4=1 ∨ S6=1 ∨ L7=1** |
| 9 | **PaM** | PeM' | M'eP | M'aP' | **V1(S3)=0, V1(S4)=0** | | | | | | |
| 10 | **PeM** | PaM' | MeP | MaP' | **V1(S1)=0, V1(S6)=0** | | | | | | |
| 11 | **PiM** | PoM' | MiP | MoP' | **V1(L2)=1** | | | | | | |
| 12 | **PoM** | PiM' | M'iP | M'oP' | **V1(L6)=1** | | | | | | |
| 13 | **P'aM'** | P'eM | MeP' | MaP | **V1(S5)=0, V1(S7)=0** | | | | | | |
| 14 | **P'eM'** | P'aM | M'eP' | M'aP | **V1(S2)=0, V1(S8)=0** | | | | | | |
| 15 | **P'iM'** | P'oM | M'iP' | M'oP | **V1(L10)=1** | | | | | | |
| 16 | **P'oM'** | P'iM | MiP' | MoP | **V1(L9)=1** | | | | | | |

## 2.2 Step II: Propagation Rules

The conditional propagation rules permit the displacement of existential values, namely 1, from points to lines and from lines to surfaces. These movements are needed for deriving all of the permissible conclusions; otherwise some of them would be missed. Figure 2 shows the general scheme of geometrical status before/after the application of the rules. The point $P$ is located in the intersection of the lines with the four adjacent surfaces. The line $L$ is located in the border of two adjacent surfaces. The Fig. 2 expresses graphically the following formal rules:

1. If $V_i(P) = 1$, $V_i(S_1) = 0$ and $V_i(S_2) = 0$ then $V_{i+1}(L) = 1$ and $V_{i+1}(P) = u$
2. If $V_i(L) = 1$ and $V_i(S_1) = 0$ then $V_{i+1}(L) = u$ and $V_{i+1}(S_2) = 1$.

First, we should apply the point-to-line rules and right after that the line-to-surface rules. The case-by-case (automated) verification of the algorithm shows that:

(i) The application of the rules is confluent and it always terminates. Hence, one always reaches to the unique result—no matter in which order the rules are applied.
(ii) No point with the value 1 stands in the intersection of four adjacent empty surfaces, i.e. surfaces with value 0.
(iii) No line with the value 1 stands between two adjacent empty surfaces, i.e. surfaces with value 0.

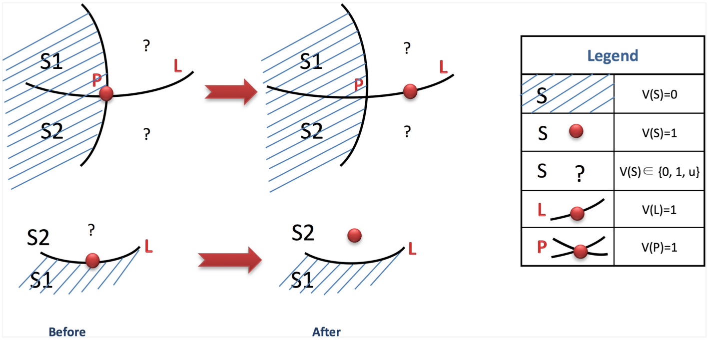


**Fig. 2.** Two general schemes of the propagation rules

### 2.3 Step III: Interpreting the Result of the Valuations

In this step, we interpret the Venn-I diagram in order to either generate all the possible permissible conclusions from two premises or to check the validity of a given syllogism. This can be considered as an inverse counterpart of the first phase. We can do this task by reading the valuations provided in the rows (17–24) of the Table 1. As one can observe, all of the valuations for the universal propositions are formulated conjunctively while the existential propositions are formulated disjunctively. Each valuation formula should be interpreted as the condition that when satisfied let us mark the relevant conclusion as derivable.

## 3 Conclusion and Possible Extensions

We have implemented and verified our proposal by means of spread-sheet programming[3]. Each point and line trigger four and two conditional propagation rules, respectively. So, the proposed procedure employs only 48 rules which are not a lot in terms of quantity comparing to $32,768 (= 32 * 32 * 32)$ cases as the large number of potential syllogistic forms. This case study and its implementation support the idea that there is no intrinsic difference between symbolic and diagrammatic reasoning as far as their underlying logics go hand in hand together. We can express that our proposed propagation rules are the implicit logic underlying Venn diagram, and with a bit of modification, we can apply it to other proposals that exist in the literature [1].

**Acknowledgement.** We would like to show our gratitude to Gholamreza Zakiani for his enlightening suggestions in the early stage of this paper.

---

[3] Please visit https://sites.google.com/view/mehdimirzapour/publications for getting access to the spreadsheet and the codes.

## References

1. Moktefi, A., Pietarinen, A.V.: On the diagrammatic representation of existential statements with Venn diagrams. J. Logic Lang. Inf. **24**(4), 361–374 (2015)
2. Moktefi, A., Shin, S.J.: A history of logic diagrams. In: Gabbay, D.M., Pelletier, F.J., Woods, J. (eds.) Logic: A History of Its Central Concepts, pp. 611–682. North-Holland, Amsterdam (2012)
3. Shin, S.J.: The Logical Status of Diagrams. Cambridge University Press, Cambridge (1994)

# Making Sense of Schopenhauer's Diagram of Good and Evil

Jens Lemanski[1(✉)] and Amirouche Moktefi[2]

[1] FernUniversität in Hagen, Hagen, Germany
jens.lemanski@fernuni-hagen.de
[2] Tallinn University of Technology, Tallinn, Estonia
amirouche.moktefi@ttu.ee

**Abstract.** It is little known that Schopenhauer (1788–1860) made thorough use of Euler diagrams in his works. One specific diagram depicts a high number of concepts in relation to Good and Evil. It is, hence, uncharacteristic as logicians of that time seldom used diagrams for more than three terms (the number demanded by syllogisms). The objective of this paper is to make sense of this diagram by explaining its function and inquiring whether it could be viewed as an early serious attempt to construct complex diagrams.

**Keywords:** Schopenhauer · Euler diagram · Diagram for *n* terms
Eristic · Persuasion

## 1 Introduction

Arthur Schopenhauer used numerous Euler diagrams in his writings and lectures. Although most of them are easily understood by the modern reader, an uncommon diagram deserves a specific attention (Fig. 1). It represents a complex network of relations between about thirty-five spheres located within a space between large areas that stand for 'Good' and 'Evil'. This diagram depicts an unusually high number of concepts in a period where diagrams were mainly used to depict relations between three terms at most. The objective of this paper is to analyze and to make sense of Schopenhauer's diagram of Good and Evil.

## 2 Schopenhauer's Diagram

Schopenhauer is seldom mentioned in history of logic literature. It is true that little logic is found in the writings that were published in his lifetime. His doctoral thesis, *On the Fourfold Root of the Principle of Sufficient Reason* (1813), contains some remarks on the structure of logic, on conceptual logic and on metalogical topics. These themes are slightly revised in Schopenhauer's main work *The World as Will and Representation* (1819) where some Euler diagrams are included. In 1844, Schopenhauer expanded the conceptual logic of the supplementary volume of the *The World* by adding a logic of judgement and a logic of syllogisms. For a much longer exposition of logic and eristic, one must consult the lesser-known Berlin lectures of the 1820s, first published in 1913

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 721–724, 2018.
https://doi.org/10.1007/978-3-319-91376-6_67

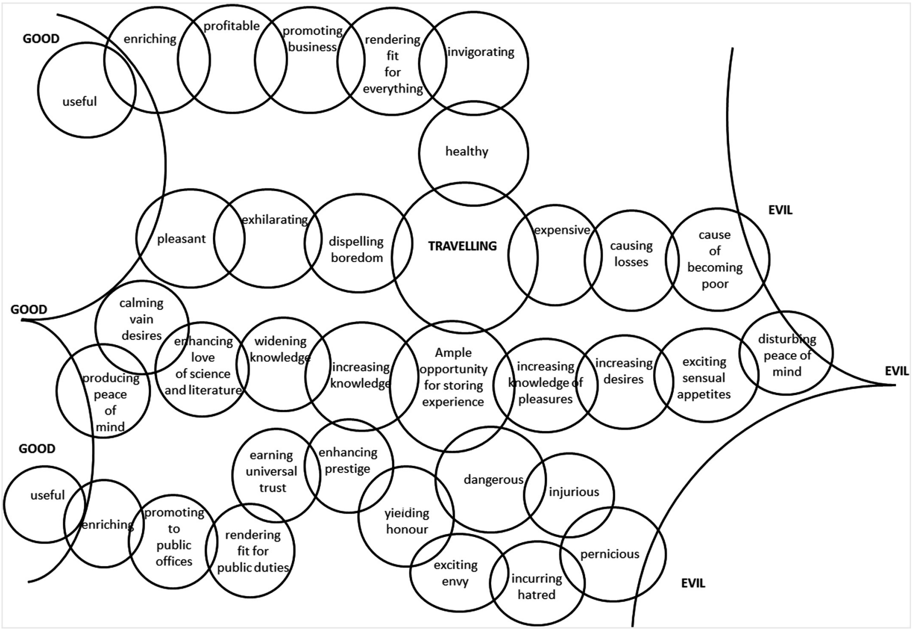


**Fig. 1.** Schopenhauer's diagram (redrawn from [3], pp. 48–49).

([1], pp. 234–366; [2]). Contrary to previous works, the lectures contain numerous logic diagrams. This contrast may be due to the nature of the writings: Schopenhauer regarded his main work as rather popular philosophical treatise, whereas the lectures were intended for students and academic philosophers. The diagram of Good and Evil is found in the section on eristic, i.e. § 9 of the *The World* ([3], pp. 48–49), but a slightly altered form also appeared in the eristic of the Berlin lectures ([1], pp. 363–366).

In Schopenhauer's logic (and eristic), every concept is said to have a sphere, i.e. an extension, and this sphere commonly stands in relation to other spheres. Schopenhauer thoroughly represented these relations between spheres with the help of circles. Although he did not think highly of the practical utility of logic, Schopenhauer certainly made high claims for logic diagrams which could be used to produce the rules on which syllogistic rests: "This schematism of concepts, which has been fairly well explained in several textbooks, can be used as the basis of the theory of judgments, as also of the whole syllogistic theory, and in this way the discussion of both becomes very easy and simple. For all the rules of this theory can be seen from it according to their origin, and can be deduced and explained" ([3], p. 44).

## 3 Diagrams for a High Number of Terms

It is well-known that the design of logic diagrams for more than three terms became prominent in the late nineteenth-century. Their need was not previously felt by the logicians who worked within the syllogistic tradition. Indeed, only three terms are found in syllogisms, and more complex problems were usually reduced to series of syllogisms. For instance, there are no higher diagrams in Euler's *Letters to a German Princess* (1768). Although complex diagrams occasionally appeared in literature, one had to wait for John Venn to find a systematic discussion of the topic. Indeed, in the footsteps of George Boole, Venn (and several of his followers) aimed at a graphical method for handling problems involving any number of terms [4].

Venn himself claimed the novelty of his work on diagrams for more than 3 terms, acknowledging only one earlier attempt: "The traditional logic has been so entirely confined to the simultaneous treatment of three terms only (this being the number demanded for the syllogism) that hardly any attempts have been made to represent diagrammatically the combinations of four terms and upwards. Almost the only serious attempt that I have seen in this way is by Bolzano" ([5], p. 511). Schopenhauer's diagram of Good and Evil clearly represents relations between numerous spheres of concepts. Interestingly, many similar diagrams for more than three terms are found in Schopenhauer's writings, although none reaches, or even approaches the complexity of the Good and Evil diagram. Hence, one might legitimately wonder if Schopenhauer's work could be considered as another 'serious attempt' (in Venn's words) to depict such complex diagrams.

## 4 The Route to Good and Evil

In order to understand the working of the diagram of Good and Evil, let us consider the sphere that stands for 'Traveling' and how it is connected to Evil. The sphere of 'Traveling' partly intersects with that of 'Expensive'. Hence, the diagram indicates that 'Some traveling is expensive'. However, a speaker might attempt to persuade that traveling is expensive, by merely considering those travels that are expensive and disregarding those that, possibly, are not. That would be incorrect for such a speaker will operate an undue generalization. However, Schopenhauer precisely argues that "the *art of persuasion* depends on our subjecting the relations of the concept-spheres to a superficial consideration only, and then determining these only from one point of view, in accordance with our intentions" ([3], p. 49). With this technique, our persuasive speaker will chain 'Traveling' to 'Expensive', then the latter to the sphere of 'Causing losses' which is itself connected to 'Cause of becoming poor'. The latter is shown to intersect with 'Evil', hence, reaching the ultimate destination of the speaker's argument. Another speaker who wishes to persuade us that 'Traveling' is rather 'Good' will follow the opposite path. He would argue that traveling is dispelling boredom and hence is exhilarating. Then, exhilaration is said to be pleasant and hence, good. Both speakers apply the same persuasive (fallacious) method which is summarized by Schopenhauer as follows: "The sphere of a concept is almost invariably shared by others, each of which contains a part of the province of the first sphere, while itself including something more besides. Of these latter concept-spheres we allow only that sphere to be elucidated under which we wish to subsume the first concept, leaving the rest unobserved, or keeping them concealed. On this trick all the arts of persuasion, all the more subtle sophisms, really depend" ([3], p. 49).

It has been argued that the persuasive strategy described by Schopenhauer in this diagram exploits the ambiguity of the concept-spheres themselves ([6], pp. 83–84). We rather argue that the strategy involves a shift in the presentation of the relations between spheres, exhibiting partial inclusion as a strict inclusion. Schopenhauer's diagram depicts 'routes' of persuasion that are used by speakers to convince of the belonging of a specific concept to Good or Evil, depending on the speaker's intention. In Schopenhauer's words: "thus one can represent the sphere of a concept A, which lies only partly in another B, but partly also in a completely different C, now after one's subjective intention as lying completely in the sphere B, or in the C" ([1], p. 364). Hence, each relation between successive concepts depicts a step in the argument. However, Schopenhauer does not attempt to depict the relation between non-successive concepts. For instance, while some traveling is said to be expensive, and expensiveness is shown to cause losses, the diagram does not depict the expected relation of (some) traveling causing losses. So, it cannot be said that Schopenhauer attempted to represent the actual relations of a high number of concepts, but merely the relations of a high number of concept considered in pairs.

## 5 Conclusion

A close examination of Fig. 1 shows that Schopenhauer does not aim at representing the whole relations between the concepts that are depicted, but rather to show 'routes' that connect each sphere to Good and Evil, at the speaker's convenience. This is an important difference between the n-term diagrams of Schopenhauer's eristic and Venn's symbolic logic: The diagram of Good and Evil is not embedded in logic, but in the eristic and is thus an early 'serious attempt' to visualize the art of persuasion.

## References

1. Schopenhauer, A.: Philosophische Vorlesungen (Sämtliche Werke IX). Edited by Deussen, P., Mockrauer, F. Piper & Co, München (1913)
2. Lemanski, J.: Logik und Eristische Dialektik. In: Koßler, M., Schubbe, D. (eds.) Schopenhauer Handbuch: Leben – Werk – Wirkung, pp. 160–169. Metzler, Stuttgart (2018)
3. Schopenhauer, A.: The World as Will and Representation, vol. I. Translated by Payne, E.F. J. Dover, New York (1969)
4. Moktefi, A., Bellucci, F., Pietarinen, A.-V.: Continuity, connectivity and regularity in spatial diagrams for N Terms. In: Burton, J., Choudhury, L. (eds.), DLAC 2013: CEUR Workshop Proceedings Diagrams, Logic and Cognition, vol. 1132, pp. 31–35 (2014)
5. Venn, J.: Symbolic Logic, 2nd edn. Macmillan, London (1894)
6. Bornedal, P.: The Surface and the Abyss: Nietzsche as Philosopher of Mind and Knowledge. De Gruyter, Berlin (2010)

# How Cross-Representational Signaling Affects Learning from Text and Picture: An Eye-Tracking Study

Juliette C. Désiron[1](✉), Mireille Bétrancourt[1], and Erica de Vries[2]

[1] University of Geneva, Geneva, Switzerland
{Juliette.Desiron,Mireille.Betrancourt}@unige.ch
[2] Univ. Grenoble Alpes, Grenoble, France
erica.devries@univ-grenoble-alpes.fr

**Abstract.** Multimedia learning research pointed out that adding a picture to a text is not systematically beneficial to learners. One of the most influential factors is the necessity for learners to identify mutually referring information in the written and pictorial representations. This study investigates how Cross-Representational Signaling (CRS) facilitates learning from multimedia document. In this study, CRS is implemented by mutually referring visual and verbal cues which highlight semantic links between text and picture. Two versions of the same multimedia document explaining the risks of being caught in a rapid, with or without CRS, are compared. The study that is still ongoing will provide data on online processing (eye-tracking data) and learning outcomes. The results will provide insights on the use of CRS to improve the design of instructional diagrams.

**Keywords:** Multimedia learning · Eye-tracking · Signaling
Cross-representational signaling

## 1 Theoretical Framework

Diagrams and pictorial representations are often used to support comprehension of instructional documents. Multimedia learning research showed that learning with multiple representations (particularly written text and pictures) can be beneficial to comprehension provided that learners can identify links between representations through cross-references [3, 10]. The most widely accepted models of multimedia learning (CTML from Mayer [6]; ITPC from Schnotz and Bannert [10]) claim that information from verbal and pictorial representations are first processed by media specific (verbal or pictorial) channels before being integrated in a coherent model of the situation relying on both those representations and previous knowledge. The latest version of the ITPC model from Schnotz [9] includes a coherence principle which predicts that "students learn better from words and pictures than from words alone if the words and pictures are semantically related to each other" ([9] p. 23), especially for students with poor reading skills or little prior knowledge.

---

The original version of this chapter was revised: The publishing mode has been changed to open access. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-91376-6_82

© The Author(s) 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 725–728, 2018.
https://doi.org/10.1007/978-3-319-91376-6_68

Effectively guiding learners' integration processes may be channelled through the insertion of visual or verbal cues in either one or both verbal and pictorial representations [11]. A meta-analysis by Richter et al. [8] found an overall significant beneficial effect of signaling text-picture relations on comprehension that was more profitable to low to medium prior-knowledge than to high prior-knowledge learners. These results confirm the ITPC [9] claim that supporting text-picture semantic links facilitates the construction of a coherent mental representation. A possible moderating effect of reading abilities was however not investigated.

Kalyuga et al. ([4] Exp. 2) used interactive colour coding in both representations to facilitate search of corresponding verbal and pictorial elements. The cueing group performed significantly better than the no-cueing group. Using eye-tracking to compare the use of verbal cues (labelling) in the pictorial representation Mason et al. [5] found that more integrative processing, measured through eye-fixations, occurred with labelled pictures. This research shows that eye-tracking data can give interesting insights in on-line processes of text-picture integration.

In the present study, we implemented Cross-Representational Signaling (CRS) through colour coding cues and picture labelling to highlight semantic links between written texts and visual pictures. Two versions (with or without CRS) of the same multimedia document (a 5-page text and picture instruction explaining the risks of being caught in a rapid) were designed. After completing reading skills tests, participants learned with one of the two versions of the multimedia document and answered comprehension (text-based and inference) questions.

We assume that CRS facilitates the construction of a coherent mental model, which should lead to better comprehension scores, especially for students with lower reading skills. Eye-tracking data will provide insights on the way CRS affects the processing of instructional diagrams. In particular, following Mason et al. [5], we expect that signaling in the text will prompt exploration of the pictures and increase the total time spent on the pictures.

## 2 Method

The experimental material was a 5-page expository document including text and static representational pictures on how to escape the Maytag effect when being caught in a rapid. The material was carefully selected and designed to ensure that both media were necessary for no prior knowledge learners to comprehend the document. The pictures were designed to be representational in the sense of Carney and Levin [1]. CRS encompassed the following mutually referencing verbal and visual cues: colours, symbols and labels (see Figs. 1 and 2). The material was presented on a 23" screen, and participants' eye-movements are recorded with a Tobii TX300.

Participants' prior knowledge was evaluated online, before the experiment, with a self-assessed multiple-choice knowledge questionnaire. Because the ITPC advocates that semantic links are helpful only to low prior knowledge readers, only participants with low or no prior knowledge on the topic were recruited. During the experiment, participants completed two reading skills assessments (a vocabulary test from Deltour [2] and an inference generation test adapted from Meteyard et al. [7]). Then they studied the multimedia document in one of two experimental conditions (with vs. without CRS). After reading, participants completed a 7 items Likert scale

[...]

Formée en ***amont*** (avant), **la lame d'eau de profondeur**, échappe toutefois au **tourbillon** pour ressurgir en ***aval*** (après) d'un **bouillonnement d'eau**. Dans celui-ci la densité de l'eau est divisée par 2, ce qui limite la flottaison.

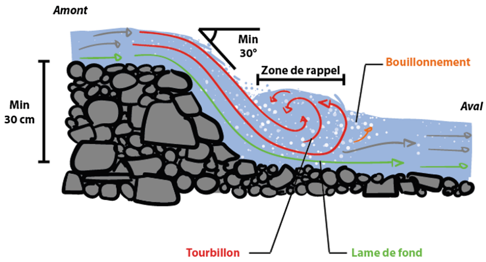


**Fig. 1.** Sample of the material with CRS

[...]

Formée en *amont* (avant), la lame d'eau de profondeur, échappe toutefois au tourbillon pour ressurgir en *aval* (après) d'un bouillonnement d'eau. Dans celui-ci la densité de l'eau est divisée par 2, ce qui limite la flottaison.

**Fig. 2.** Sample of the material without CRS

questionnaire on motivation, perceived difficulty and perceived effort. They ended with the comprehension test, with 13 open-ended questions at three levels: text-base comprehension, local bridging and global-bridging inference. A drawing task was also included, in which participants had to draw and name the different currents involved in the formation of a whitewater.

The experiment was still running when we wrote this paper. A random sample of 40 to 50 undergraduate university students in education sciences or psychology will be recruited overall.

## 3 Data Analyses and Expected Results

Following previous research in multimedia learning using eye-tracking as an online measure of comprehension [3, 5], we will analyze the collected data with first-pass and second-pass fixations. Specifically, we will consider fixations as gazes and focus on look from text to picture, both in general and with targeted AOI.

First, following the ITPC model [9] we expect that multimedia comprehension will be higher in the CRS than in the control condition, especially for students with low reading skills. Regarding on-line processing, we expect that participants reading the multimedia document with CRS will look at the picture during first-pass and second-pass reading more often than participants without CRS. Indeed, research by Mason et al. [5] pointed out that a picture with verbal cues elicited more integration with the text than a picture without verbal cues. Further exploratory analyses of eye-tracking data will provide insights on how text-picture integration processes differ with and without CRL. Participants reading skills will be inserted in the analyses as a potential moderator.

This study will contribute to test an implementation of the *coherence* condition, theoretically developed in the ITPC model, when a document is designed with

Cross-Representational Signaling. The findings will provide guidelines regarding the design of commented diagrams used for instructional or public awareness purposes.

**Acknowledgments.** This study was supported by the Swiss National Science Foundation with a Doc.CH grant attributed to the first author [P0GEP1_165256].

## References

1. Carney, R.N., Levin, J.R.: Pictorial illustrations still improve students' learning from text. Educ. Psychol. Rev. **14**(1), 5–26 (2002). https://doi.org/10.1023/A:1013176309260
2. Deltour, J.J.: Echelle de vocabulaire de Mill Hill de JC Raven: Adaptation française. In Manuel des Raven. Braine le Château. Oxford Psychologists Press, Belgique (1993)
3. Hegarty, M., Just, M.A.: Constructing mental models of machines from text and diagrams. J. Mem. Lang. **32**, 717–742 (1993). https://doi.org/10.1006/jmla.1993.1036
4. Kalyuga, S., Chandler, P., Sweller, J.: Managing split-attention and redundancy in multimedia instruction. Appl. Cogn. Psychol. **25**(Suppl. 1), 351–371 (1999). https://doi.org/10.1002/acp.1773
5. Mason, L., Pluchino, P., Tornatora, M.C.: Effects of picture labeling on science text processing and learning: evidence from eye movements. Read. Res. Q. **48**, 199–214 (2013). https://doi.org/10.1002/rrq.41
6. Mayer, R.E.: Cognitive theory of multimedia learning. In: Mayer, R.E. (ed.) The Cambridge Handbook of Multimedia Learning, pp. 31–46. Cambridge University Press, Cambridge (2005). https://doi.org/10.1017/CBO9780511816819.004
7. Meteyard, L., Bruce, C., Edmundson, A., Oakhill, J.: Profiling text comprehension impairments in aphasia. Aphasiology **29**(1), 1–28 (2015). https://doi.org/10.1080/02687038.2014.955388
8. Richter, J., Scheiter, K., Eitel, A.: Signaling text-picture relations in multimedia learning: a comprehensive meta-analysis. Educ. Res. Rev. **17**, 19–36 (2016). https://doi.org/10.1016/j.edurev.2015.12.003
9. Schnotz, W.: Integrated model of text and picture comprehension. In: Mayer, R.E. (ed.) The Cambridge Handbook of Multimedia Learning, 2nd edn., pp. 72–103. Cambridge University Press, New York (2014). https://doi.org/10.1017/cbo9780511816819.005
10. Schnotz, W., Bannert, M.: Construction and interference in learning from multiple representation. Learn. Instr. **13**, 141–156 (2003). https://doi.org/10.1016/S0959-4752(02)00017-8
11. van Gog, T.: The signaling (or cueing) principle in multimedia learning. In: Mayer, R.E. (ed.) The Cambridge Handbook of Multimedia Learning, 2nd edn., pp. 263–278. Cambridge University Press (2014). https://doi.org/10.1017/cbo9781139547369.014

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, a link is provided to the Creative Commons license and any changes made are indicated.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

# Effect of Handedness on Mental Rotation

Gunta Krumina[1(✉)], Jurgis Skilters[1], Annija Gulbe[1],
and Vsevolod Lyakhovetskii[2]

[1] University of Latvia, Riga, Latvia
gunta.krumina@lu.lv
[2] Pavlov Institute of Physiology, Russian Academy of Sciences,
St.Petersburg, Russia

**Abstract.** The impact of the dominant hand on the response time and precision in mental rotational tasks seems to be controversial. The goal of this study was to compare the differences in response times of mental rotation tasks when the task is performed with the dominant or non-dominant hand. In this study, 44 right-handers and 45 left-handers participated in mental rotation tests with 2-D and 3-D figures. Findings indicate that the right-handers had shorter response times than left-handers in tests with both types of figures.

**Keywords:** Reaction time · 2-D figures · 3-D figures · Right-handers
Left-handers

## 1 Introduction

Spatial visualization and mental rotation are considered to be among the at the core of human visuo-spatial abilities. Shepard and Metzler [1] and Cooper [2] provided initial results indicating that humans have a measurable ability to mentally rotate two-dimensional or three-dimensional objects. Further, they provided plausible Reaction-Time-based methods for deciding whether these figures are the same objects in different positions.

Several studies emphasize the importance of (1) strategy, (2) impact of gender, (3) age, and (4) the properties of the rotational stimulus in the visual field. Many of the studies have been conducted with right-handed subjects; however, almost no studies have explored the impacts of handedness on the rotational tasks. Right-handers and left-handers are described as people who tend to use the right or the left hand as the dominating and are convenient for doing different tasks. The impact of the dominant hand on the response time and precision in mental rotational tasks seems to be controversial though. Additionally, studies by Metzler and Shepard [3] and Jones and Anuza [4] indicate that left-handers are less effective than right-handers in mental rotation tasks.

However, Herrmann and van Dyke [5] argues that the opposite is the case: that mental rotation is performed more quickly by left-handed participants (with their dominant hand) and 2-dimensional images are retained by the linear relationship between the reaction time and the turning angle. The study by Peters et al. [6] does not show any differences in mental rotation between right- and left-handers, but only that

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 729–733, 2018.
https://doi.org/10.1007/978-3-319-91376-6_69

there are differences between male and female subjects according to their responses. One of the explanations for the handedness differences is based on the specialization of the cerebral hemispheres; verbal or symbolic ability is considered to be provided by the left hemisphere of the brain, while the right hemisphere is responsible for processing spatial information [7]. All these studies are considered to be inconclusive [6] as it can be assumed that the process of spatial encoding for the left and right hand varies according to the hand they use for responding [8].

The present study aimed to investigate the possible advantages of right- and left-handers in the tasks of mental rotation. The goal of the study was to compare the differences in response times when performing mental rotation tasks with the dominant or non-dominant hand.

## 2 Method

Mental rotation tasks were displayed on a computer screen at a distance of 60–65 cm from the subjects. Eighty-nine (89) subjects participated in the study (44 right-handers and 45 left-handers) and the age range was 19–33. Subjects without a dominant hand were excluded from the study. The Right-hand test (R-test) was performed by 21 right-handed and 21 left-handed subjects; the Left-hand test (L-test) was performed by 23 right-handed and 24 left-handed subjects. The 2-D figures used in the mental rotation tests and stimuli used were adapted from Cooper [2] and the 3-D figures from Shepard and Metzler [1]. The computer program presented 2-D or 3-D figures in a random order at four different angles (0°, 60°, 120° and 180°). 2-D figures chosen were those that were closer to recognizable objects, such as fish, flower, dog (see Fig. 1 2-D stimuli), but 3-D stimuli chosen were abstract objects.

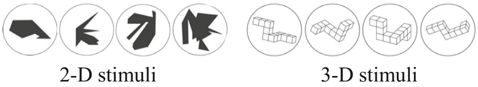


**Fig. 1.** The 2-D and 3-D figures used in the study. Each figure was shown three times and presented in four different angles of rotation.

Subjects sat in front of the computer screen and positioned their right and left index fingers on a keyboard button. The group of right-handers (RH) performed the Right-hand test (R-tests - RH) and gave the answers about the similar figures displayed on the screen with the right hand and about the different figures – with their left hand. The second right-hander group performed the Left-hand test (L-test - RH) and reacted to similar figures by pressing the button with the left hand. The same procedure was used for the two groups of left-handers. The test was not done for the same subjects in both conditions because of concerns about learning effects. All figures (2-D and 3-D) were randomly presented three times in one test session. The program recorded the response times for each figure and the correct and incorrect answers given.

In the data analysis, non-parametric statistical methods were used to compare response times. The Mann-Whitney U test was used to compare the mean response times of all correctly recognized figure pairs and differences in results between the right- and left-handers. Non-parametric one-way analysis using the Kruskal-Wallis test was used to compare the difference between the four orientations (0°, 60°, 120°, and 180°).

## 3 Results and Discussion

The Kruskal-Wallis test of variance showed significant effect ($p < 0.0001$): in all tests (R- and L-tests with 2-D and 3-D figures) the response time increased as a function of the rotation angle. According to these results, it can be seen that there is a significant impact of test type on the performance of the right-handers. The right-handers who performed the test with their right-hand showed shorter response times than when they performed the test with their left-hand ($p < 0.05$) for both tasks, (2-D and the 3-D figures) (see Fig. 2, for RH results). In contrast, the difference between right- or left-hand tasks with 2-D and 3-D figures was not observed ($p > 0.3$) in the left-handers, (see Fig. 2, for LH results).

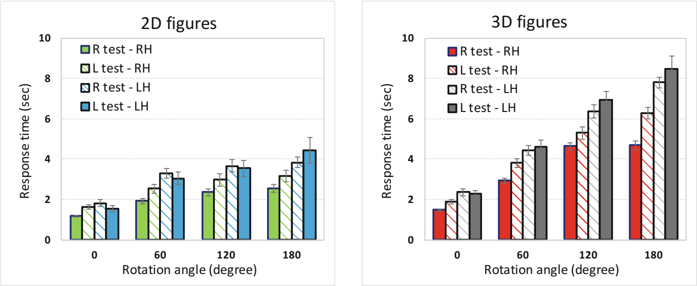


**Fig. 2.** Right-handers (RH) and left-handers (LH) response times when using the dominant and non-dominant hand. The graphs present mean ± SE results.

Response times in R-tests for left-handers were crucially different when compared with the right-handers. The right-handers rotated the 2-D and 3-D figures ($p < 0.0001$) faster than the left-handers. However, results of L-tests showed no statistically significant difference ($p > 0.3$) between the subjects with the left dominant hand and the subjects with the right dominant hand. Test results achieved by both groups using their dominant hand showed that the left-handers' response times were longer than the right-handers' response times ($p < 0.01$). When using their non-dominant hand, right-handers mentally rotated 2-D figures faster than left-handers ($p = 0.04$) but response times for 3-D figures rotation showed no significant difference ($p = 0.07$). Left-handers performed the mental rotation tasks slower in all cases.

A statistically significant difference between right-handers and left-handers can be observed when using their dominant hand; the left-handed subjects performing the tasks slower than the right-handers ($p < 0.05$). However, if the impact of non-dominant hand is taken into account, results show no significant differences between the groups. At the same time, left-handers show longer reaction times in the 3-D test in all test conditions. These data indicate that right-handers have shorter response times than left-handers in all test conditions. This may be due to many reasons but most likely there is some sensorimotor overload which impacts on those tendentially larger reaction times in left-handers.

The results from this study have relevance for diagram research since mental rotation task occurs frequently while reading information in diagrams. This work shows some tentative evidence that right-handers and left-handers perceive rotation tasks differently. This has important consequences for information visualization and should be kept in mind when diagrams containing mental rotation are used. According to these results, it is possible to hypothesize that the underlying process of using visualizations by offloading cognition on external visuo-spatial representations can be impaired in left-handers. Although more research on rotational diagrams is needed, these results relate to the principles of perception and expressiveness of displays [9, 10], suggesting that left-handers' capacity limitations may impact on the results of mental rotation.

## 4 Conclusion

From these results it can be concluded that right-handers have shorter response times than left-handers in mental rotation of 2-D and 3-D figures. These results also indicate that the impact of handedness is more complex than one might assume as results from the group of left-handers do not indicate a linear correlation between the rotation angle and the response time in the 2-D version of the task. A general observation from these results can be made that sensorimotor task complexity for left-handers increases the reaction times of this group and this should be taken into account in the development of visual materials.

## References

1. Shepard, R.N., Metzler, J.: Mental rotation of tree-dimensional objects. Science **171**(3972), 701–703 (1971)
2. Cooper, L.A.: Mental rotation of random two-dimensional shapes. Cogn. Psychol. **7**(1), 20–43 (1975)
3. Metzler, J., Shepard, R.N.: Transformational studies of the internal representation of three-dimensional objects. In: Solso, R.L. (ed.) Theories in Cognitive Psychology: The Loyola Symposium. Lawrence Erlbaum, Oxford (1974)
4. Jones, B., Anuza, T.: Effects of sex, handedness, stimulus and visual field on "mental rotation". Cortex **18**(4), 501–514 (1982)
5. Herrmann, D.J., van Dyke, K.A.: Handedness and the mental rotation of perceived patterns. Cortex **14**(4), 521–529 (1978)
6. Peters, M., Reimers, S., Manning, J.T.: Hand preference for writing and associations with selected demographic and behavioural variables in 255,100 subjects: the BBC internet study. Brain Cogn. **62**(2), 177–189 (2006)

7. Levy, J., Reid, M.: Variations in cerebral organization as a function of handedness, hand posture in writing, and sex. J. Exp. Psychol. Gen. **107**(2), 119–144 (1978)
8. Arend, I., Weiss, P.H., Timpert, D.C., Fink, G.R., Henik, A.: Spatial coding as a function of handedness and responding hand: Theoretical and methodological implications. PLoS ONE **11**(3), e0151979 (2016)
9. Hegarty, M.: The cognitive science of visual-spatial displays: implications for design. Top. Cogn. Sci. **3**(3), 446–447 (2011)
10. Tversky, B.: Visualizing thought. Top. Cogn. Sci. **3**(3), 499–535 (2011)

# Euler Box Diagrams to Represent Independent and Non-independent Events

James E. Corter(✉)

Teachers College, Columbia University, New York, NY 10027, USA
`jec34@tc.columbia.edu`

**Abstract.** Venn and Euler diagrams are valuable tools for representing the logical set relationships among events. Proportional Euler diagrams add the constraint that the areas of diagram regions denoting various compound and simple events must be proportional to the actual probabilities of these events. Such proportional Euler diagrams allow human users to visually estimate and reason about the probabilistic dependencies among the depicted events. The present paper focuses on the use of proportional Euler diagrams composed of rectangular regions and proposes an enhanced display format for such diagrams, dubbed "Euler boxes", that facilitates quick visual determination of the independence or non-independence of two events and their complements. It is suggested to have useful applications in exploratory data analysis and in statistics education, where it may facilitate intuitive understanding of the notion of independence.

**Keywords:** Euler diagrams · Proportional Euler diagrams · Probability

## 1 Introduction

The use of closed-curve diagrams to represent sets is usually attributed to Leonhard Euler. This work was expanded and applied to the study of logic by John Venn. In recent decades, such diagrams have become widely used in logic and mathematics education [1, 2]. The diagrams have important applications in probability theory and statistics education [3–5]. In some scientific domains, *proportional* Euler diagrams have been used as a means to visualize data [6, 7]. Proportional (or "scaled") Euler diagrams impose the constraint that the distinguished subareas in the diagram should be proportional to a set of weights. Most often, the weights represent the probability or relative frequency of the corresponding event. Recent work in computer and information science has developed algorithms for automatic generation of proportional Venn and Euler diagrams, using both circles and polygons [8–12]. Some of these algorithms are widely available as R packages, among them *venneuler*, *VennDiagram*, and *eulerr*.

The present paper proposes an enhanced display format for proportional Euler diagrams composed of rectangular regions ("Euler boxes"), that represents the areas of simple and joint events in a way that facilitates user inferences about the independence of simple or marginal events. The diagram is advocated as a way for students of probability theory to visualize and better understand the concept of the independence of two events or random variables, and as a form of graphical data analysis providing a means to visually detect non-independence in contingency tables.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 734–738, 2018.
https://doi.org/10.1007/978-3-319-91376-6_70

# 2 The Euler Box Diagram

## 2.1 Example: Independent Events

The Euler box diagram is a type of proportional Euler diagram using rectangular regions. Such diagrams can be seen as direct visualizations of a contingency table [8]. Consider an example of independent events generated by the flip of a coin, resulting in Heads (H) or Tails (T), and the simultaneous roll of a die, with 4 of the faces colored Red and 2 faces White. The joint event probabilities are easily calculated, and can be represented by the diagram shown in Fig. 1.

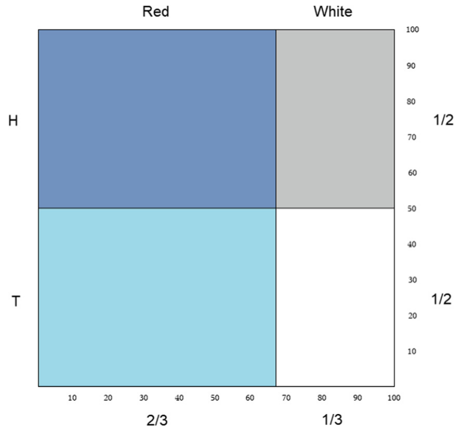


**Fig. 1.** Euler box diagram of independent events

In this diagram, the overall area is assumed to be 1, the total probability of the outcome space. Probabilities of the simple events Red and White are represented by the overall areas of the regions indicated by the blue and white colors, while the probability of Heads (H) is depicted by the overall area with grey shading. Texture shading could also be used to differentiate one facet of the diagram, if colors are not available. Importantly, the areas of all joint events (e.g., Heads ∩ White) are also proportional to their area. Finally, the projections of the rectangles on the axes are another way to visualize the probabilities of the simple events, as the proportions of their extent on each axis. In Fig. 1 these marginal probabilities are explicitly noted; e.g., $\mathrm{P(H)} = \frac{1}{2}$.

The diagram in Fig. 1 illustrates the concept of independence in a very direct way. One definition of independence states that two events A and B are independent if P(A ∩ B) = P(A)P(B). It is clear from the diagram that the area corresponding to (Heads ∩ Red) is simply equal to P(H)P(Red) = (1/2)(2/3) = 2/6. Another definition of independence states that two events A and B are independent if P(A|B) = P(A). It is apparent from the diagram that P(Red|H) = P(Red|T) = P(Red) = 2/3, although this involves some visual reasoning, namely mental estimation of proportional areas, e.g., the proportion of the grey-shaded region (H) that is also blue (Red), which is the visual counterpart of P(Red|H). Thus, the diagram directly represents the fundamental relationships involved in the concept of independence, which may be helpful for statistics learners.

## 2.2 Example: Euler Boxes for Non-independent Events

According to the U. S. Centers for Disease Control (CDC), smoking is a prevalent health risk for adults in the U.S. This risk is heightened for adults living in poverty: according to CDC data on smoking trends 2005–2015, in one study the rate of smoking among adults living below the federal poverty line was 32.4%; among adults at or above the poverty line it was 20.0%. According to the U.S. Census, 14.8% of U.S. adults were living below the poverty line in that period, implying the following 2 × 2 joint probability distribution (Table 1). From the table, it is difficult to tell at a glance if independence is satisfied or not.

**Table 1.** Smoking among US adults by poverty level (CDC 2005–2015): observed relative frequencies, expressed as percentages.

| | Poverty | At/above poverty | Marginal |
|---|---|---|---|
| S | 4.8% | 17.0% | 21.8% |
| NS | 10.0% | 68.2% | 78.2% |
| | 14.8% | 85.2% | |

The Euler box representing these non-independent events is constructed as follows. First, the Euler diagram representing independence of the simple or marginal events is constructed. Then, the rectangular regions representing the joint events are enlarged or reduced, so that the area of each rectangle is proportional to the observed relative frequency for that joint event, as shown in Fig. 2. This involves enlarging or shrinking each rectangular region by a factor s = $\sqrt{o_{ij}/e_{ij}}$, where $o_{ij}$ and $e_{ij}$ are the observed and expected relative frequencies for cell $(i, j)$. This scaling factor is applied to both height and width dimensions. Note that the outline of the original diagram representing independence is retained as a frame or "graphical overlay" [13] (with the same total area as the four rectangles representing the joint events). This enables direct visual

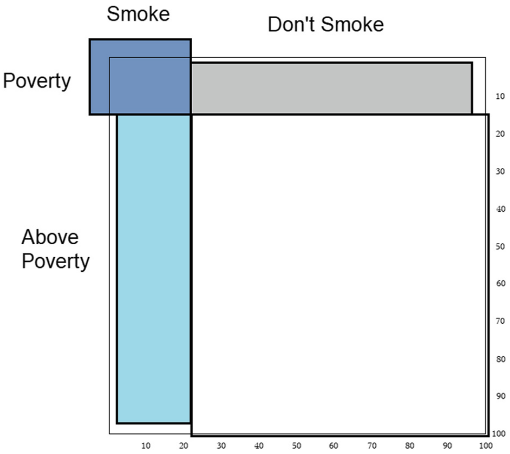


**Fig. 2.** Euler box diagram for data linking smoking and poverty.

inferences regarding whether a given cell is more or less likely than expected under independence. The Euler box in Fig. 2 visually demonstrates that the two random variables (and all joint events) are not independent, and that (for example) adults living in poverty are much more likely to smoke.

## 3 Future Directions

Future research is needed to assess the utility of the present work. In statistics education, studies might assess if the Euler box diagram indeed helps students learning probability to better understand independence. In visual data analysis, the effectiveness of the display for rapid assessment of independence relations might be assessed. Just as importantly, more work is needed to generalize the enhanced Euler boxes display format beyond the case of two dichotomous events and variables. If one (or both) of the categorical random variables has three or more levels, what adjustments or accommodations should be made to the diagram to preserve the user benefits discussed here? Can useful extensions to the case of three or more variables be devised? Future work is planned to examine these questions.

## References

1. Presmeg, N.: Research on visualization in learning and teaching mathematics. In: Gutiérrez, Á., Boero, P. (eds.) Handbook of Research on the Psychology of Mathematics Education: Past, Present and Future, pp. 205–236. Sense, Rotterdam (2006)
2. Uesaka, Y., Manalo, E., Ichikawa, S.: What kinds of perceptions and daily learning behaviors promote students' use of diagrams in mathematics problem solving? Learn. Instr. **17**(3), 322–335 (2007)
3. Corter, J.E., Zahner, D.C.: Use of external visual representations in probability problem solving. Stat. Educ. Res. J. **6**(1), 22–50 (2007)
4. Zahner, D., Corter, J.E.: The process of probability problem solving: use of external visual representations. Math. Think. Learn. **12**(2), 177–204 (2010)
5. Cheng, P.C.-H.: Probably good diagrams for learning: representational epistemic re-codification of probability theory. Top. Cogn. Sci. **3**(3), 475–498 (2011)
6. Kestler, H., Müller, A., Kraus, J., Buchholz, M., Gress, T., Liu, H., Kane, D., Zeeberg, B., Weinstein, J.: VennMaster: area-proportional Euler diagrams for functional GO analysis of microarrays. BMC Bioinform. **9**, 67 (2008)
7. Edwards, A.W.F., Edwards, J.H.: Metrical Venn diagrams. Ann. Hum. Genet. **56**, 71–75 (1992)
8. Stapleton, G., Zhang, L., Howse, J., Rodgers, P.: Drawing Euler diagrams with circles: the theory of piercings. IEEE Trans. Vis. Comput. Graph. **17**(7), 1020–1032 (2011)
9. Chen, H., Boutro, P.C.: VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinform. **12**(3), 1 (2011)
10. Wilkinson, L.: Exact and approximate area-proportional circular Venn and Euler diagrams. IEEE Trans. Vis. Comput. Graph. **18**(2), 321–331 (2012)
11. Rodgers, P., Stapleton, G., Flower, J., Howse, J.: Drawing area-proportional Euler diagrams representing up to three sets. IEEE Trans. Vis. Comput. Graph. **20**(1), 56–69 (2014)

12. Chow, S., Ruskey, F.: Drawing area-proportional venn and euler diagrams. In: Liotta, G. (ed.) GD 2003. LNCS, vol. 2912, pp. 466–477. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24595-7_44
13. Kong, N., Agrawala, M.: Graphical overlays: using layered elements to aid chart reading. IEEE Trans. Vis. Comput. Graph. **18**(12), 2631–2638 (2012)

# A Symmetry Metric for Graphs and Line Diagrams

Roman Klapaukh[1,2(✉)], Stuart Marshall[2], and David Pearce[2]

[1] Research IT Services, University College London,
London NW1 2DN, UK
`r.klapaukh@ucl.ac.uk`
[2] School of Engineering and Computer Science, Victoria University of Wellington,
Wellington, New Zealand

**Abstract.** Symmetry is often considered a desirable feature of diagrams. However, quantifying the exact amount of symmetry present is often difficult. We propose a novel symmetry metric that can score the amount of rotational, translational, and reflective symmetry present in a graph or line diagram.

**Keywords:** Symmetry detection · Line diagrams
Diagram Evaluation

## 1 Introduction

Symmetry is a property of visual layouts that is frequently considered to be desirable. Many believe greater symmetry improves understandability and that, for example, force-directed layout promotes symmetry [1,2]. In the study of Kieffer *et al.*, human subjects showed a preference for graphs with reflectional symmetry [3]. Likewise, Marriot *et al.* confirmed that various layout features—including symmetry—make graphs more memorable [4].

For graphs, Purchase defined an algorithm for computing a symmetry score and used this to test various claims [5]. Purchase's symmetry algorithm has two important limitations which this paper attempts to address. First it focused only on the symmetries of vertices, but ignored edges. Second it measured only reflective symmetries, ignoring rotational and translational symmetries.

In this paper we develop a new symmetry metric for straight line diagrams. Our symmetry metric is an extension of that described by Loy and Eklundh [6] which extracts reflective, rotational and translational symmetries of feature points from photographs. We adapt their algorithm to work on known vector lines, rather than feature points detected in raster images. Unlike points, lines can be symmetrical with respect to themselves, and for the line drawing we consider you have perfect knowledge about the lines. This holds true for images stored in vector formats (e.g. SVG).

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 739–742, 2018.
https://doi.org/10.1007/978-3-319-91376-6_71

## 2 Background

Purchase [5] defined a symmetry metric for graphs that measured how much reflective symmetry was present between vertices. However, an important limitation of Purchase's algorithm is that it does not consider edges when calculating a symmetry score. In all connected graphs which are not trees, edges will outnumber vertices. Additionally, edges are drawn as lines connecting the vertices. This results in the edges having a lot of influence on visual symmetry within the graph. Therefore, in this work we focus on quantifying the symmetry of edges.

We extend the symmetry detection method of Loy and Eklundh as it is works in 2D and can identify rotational, translational, and reflective symmetries [6]. However, Loy and Eklundh focus on identification of symmetry for points that have been extract from photographs, in order to identify regions of symmetry. However, as we consider lines, which are symmetrical with respect to themselves, we develop a new method following their process.

An important feature of Loy and Eklundh's algorithm is that it can detect when multiple different detected symmetry axes/centres are actually very similar (i.e. rather than only when they are identical), and groups them together into a single axis. This is very useful in generated diagrams as they may have minor imperfections which should not be penalised.

## 3 Symmetry Metric

Our extended symmetry metric is described in Algorithm 1. The rest of this section describes each part of the algorithm. The implementation can be found as part of our open source graph analysis library [7].

The general idea of the algorithm is that it finds every possible axis of mirror, translational, or rotational symmetry that is present in the diagram. It then votes to identify which axes affect the largest number of lines. The number of axes detected is a user defined parameter $N$. Note that the score for each type of symmetry is computed separately.

We start with an empty list of pairs of an axis (two floating point numbers) and its quality score (a number between 0 and 1). Each kind of symmetry has a different sort of axis: rotational symmetry has a centre of rotation, translational symmetry has a direction vector, and reflective symmetry has the Hough transform [8] of the mirror line.

We then convert the lines into a standard format for easy manipulation. Following Loy and Eklundh [6] we convert them into Scale Invariant Feature Transform (SIFT) [9] features. Each SIFT feature is a four-tuple consisting of a location (centre of the line), orientation (angle between the line and x-axis in degrees), scale (length of the line in pixels), and identifying characteristics (always 1).

The first set of axes we generate are those that are symmetries of an axis with itself. For reflective symmetry there are two axes that can be generated from a single line: its perpendicular bisector, and the line itself. For rotational symmetry there is one: the position of the line, as a line can be spun around

**Algorithm 1.** Symmetry metric.

```
Data: symmetryType ∈ {reflective,rotational,translational} & $N$ the number of
      axes to find
Result: score ∈ [0..1]
axes = empty list
features = Convert all edges to SIFT features
for $f_i \in$ features do
    if symmetryType == reflective then
        axes.add(perpendicularBisector($f_i$), 1)
        axes.add(parallelAxis($f_i$), 1)
    else if symmetryType == rotational then
        axes.add(getLocation($f_i$), 1)
foreach $f_i, f_j \in$ features do
    axis = find symmetry axis(symmetryType, $f_i$, $f_j$)
    quality = find symmetry quality(symmetryType, $f_i$, $f_j$)
    axes.add(axis, quality)
axes = quantiseAxes(axes)
bestAxes = pickBest(axes, N)
score = score(bestAxes)
return score
```

its centre 180° to get the same line. There are no such axes for translational symmetry. In each the symmetry is perfect, therefore the quality score for all these axes is 1. Note that these single feature axes of symmetry are not present in Loy and Eklundh's algorithm.

All other possible axes of symmetry can be generated by calculating the axes of symmetry between every pair of lines. The quality score for each axes is the product of its scale quality ($S_{ij}$) and orientation quality ($\Phi_{ij}$) scores. Each score is bound by $[0 \dots 1]$.

The scale quality ($S_{ij}$) is the same for all symmetry types and is the same as Loy and Eklundh's original paper [6]. In the equation $S_{ij}$ is the scale similarity, $s_k$ is the length of line $k$, and $\sigma_s$ is a scaling factor (sensitivity).

$$S_{ij} = \left( e^{\frac{-|s_i - s_j|}{\sigma_s (s_i + s_j)}} \right)^2 \tag{1}$$

The axis of reflective symmetry is the perpendicular bisector of the line between the line centres. The orientation quality ($\Phi_{ij}$), is adapted from Reisfeld et al. [10], with consideration for lines being symmetrical after 180° rotations.

$$\Phi_{ij} = |\cos(\theta_i + \theta_j - 2 * \theta_{ij})| \tag{2}$$

For translational symmetry the required translation is the vector difference in the position of the features. This needs to be normalised (multiplied by $-1$ if dy $<0$) to compensate for ordering. For translational symmetry the orientation quality has to be adjusted as the lines are not mirrored.

$$\Phi_{ij} = \left|\cos\left(\theta_i - \theta_j\right)\right| \tag{3}$$

For each pair of features there can be up to two centres of rotational symmetry. This is because there are two ways to line up the feature orientations, head to head and head to tail, each of which may require its own centre. When $\theta_1 = \theta_2 = 180$ both centres will be the same. This is different from the original paper where there was only one possible centre, and is a result of a line being indistinguishable after a 180° rotation. The orientation metric score is always 1.

Having enumerated all of the axes of symmetry, the quality scores are now used to vote to find the $N$ best axes. We quantise the space, to deal with minor deviations in location, and sum the symmetry scores for each distinct axis. The $N$ axes with the highest total scores are the chosen axes.

The final stage of the algorithm is to turn the set of $N$ best axes found into a number that can be used as a metric. We use the following equation:

$$\left(\frac{\sum_{axes} \text{number of lines that voted for this axis}}{N \times \text{number of lines}}\right)$$

## 4 Conclusion

We developed a novel metric to evaluate how symmetrical a given line diagram is with respect to reflective, rotational and translational symmetries.

## References

1. Eades, P.: A heuristic for graph drawing. Congr. Numer. **42**, 149–160 (1984)
2. Purchase, H.C., Cohen, R.F., James, M.: Validating graph drawing aesthetics. In: Brandenburg, F.J. (ed.) GD 1995. LNCS, vol. 1027, pp. 435–446. Springer, Heidelberg (1996). https://doi.org/10.1007/BFb0021827
3. Kieffer, S., Dwyer, T., Marriott, K., Wybrow, M.: HOLA: human-like orthogonal network layout. IEEE TVCG **22**(1), 349–358 (2016)
4. Marriott, K., Purchase, H., Wybrow, M., Goncu, C.: Memorability of visual features in network diagrams. IEEE TVCG **18**(12), 2477–2485 (2012)
5. Purchase, H.C.: Metrics for graph drawing aesthetics. JVLC **13**(5), 501–516 (2002)
6. Loy, G., Eklundh, J.-O.: Detecting symmetry and symmetric constellations of features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006 Part II. LNCS, vol. 3952, pp. 508–521. Springer, Heidelberg (2006). https://doi.org/10.1007/11744047_39
7. Klapaukh, R.: GraphAnalyser. https://github.com/klapaukh/GraphAnalyser. Accessed 20 Dec 2013
8. Duda, R.O., Hart, P.E.: Use of the Hough transformation to detect lines and curves in pictures. Commun. ACM **15**(1), 11–15 (1972)
9. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceeding of Computer Vision, pp. 1150–1157 (1999)
10. Reisfeld, D., Wolfson, H., Yeshurun, Y.: Context-free attentional operators: the generalized symmetry transform. Int. J. Comput. Vis. **14**(2), 119–130 (1995)

# Wrapping Layered Graphs

Ulf Rüegg(✉) and Reinhard von Hanxleden

Department of Computer Science, Kiel University, Kiel, Germany
{uru,rvh}@informatik.uni-kiel.de

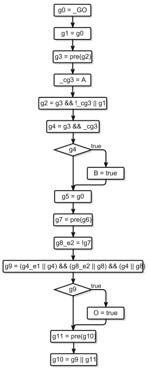
**Abstract.** We present additions to the widely-used layout method for directed acyclic graphs of Sugiyama et al. that allow to better utilize a prescribed drawing area. The method itself partitions the graph's nodes into layers. When drawing from top to bottom, the number of layers directly impacts the height of a resulting drawing and is bound from below by the graph's longest path. As a consequence, the drawings of certain graphs are significantly taller than wide, making it hard to properly display them on a medium such as a computer screen without scaling the graph's elements down to illegibility. We address this with the Wrapping Layered Graphs Problem (WLGP), which seeks for cut indices that split a given layering into chunks that are drawn side-by-side with a preferably small number of edges wrapping backwards. Our experience and a quantitative evaluation indicate that the proposed wrapping allows an improved presentation of narrow graphs, which occur frequently in practice and of which the internal compiler representation SCG is one example.

## 1 Introduction

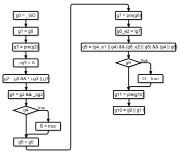
The *layer-based layout approach* is a successful method to lay out directed acyclic graphs. In its original form it has been presented by Sugiyama et al. in 1981 [4] and since then has been refined for many use cases. Healy and Nikolov give a comprehensive survey of the state of the literature in 2013 [1]. The approach's central idea is to assign nodes to indexed *layers* such that edges point from layers with lower index to layers with higher index. The overall layout task is split into three consecutive phases: *layering*, *crossing minimization*, and *coordinate assignment*. To allow cyclic graphs, an additional initial *cycle breaking* phase can be added to make the graph acyclic, and a final *edge routing* phase can be added to support different edge routing styles.

If a graph is laid out from top to bottom, the number of used layers directly impacts the height of the resulting drawing and is bound from below by a graph's longest path. Thus, an inherent problem of the layering process is that graphs with long longest paths may yield drawings which are significantly taller than wide, resulting in unfortunate aspect ratios. Certain graph types occurring in practice encourage such drawings by nature. See, for instance, Fig. 1a, which shows the drawing of a sequentialized *Sequentially Constructive Graph (SCG)* [5], a type of control flow diagram, where each node represents an execution step

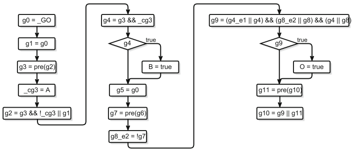
© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 743–747, 2018.
https://doi.org/10.1007/978-3-319-91376-6_72



(a) No cuts, one chunk

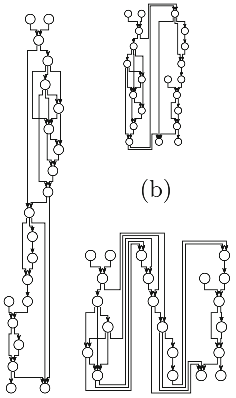


(b) One cut, two chunks



(c) Two cuts, three chunks

**Fig. 1.** A sequentialized SCG drawn with different numbers of cuts. The aspect ratios of the drawings vary significantly, each well-suited for a certain prescribed drawing area.



(b)

(a) (c)

**Fig. 2.** Three drawings of the same graph, each time with a different number of cuts and backward wrapping edges.

of a program. With increasing program sizes, either only a small part of the created drawings can be displayed on a monitor, or the drawings have to be scaled down to illegibility. However, in particular the labels within the nodes require a presentation in which they can be properly read by a user.

We address this with the *Wrapping Layered Graphs Problem (WLGP)* as defined below. It seeks for *cut indices* that split a given layering into *chunks* that are to be drawn side-by-side such that a prescribed drawing area is used as good as possible and such that a preferably small number of edges wraps backwards. Results can be seen in Figs. 1b, c and 2b, c. Note that while the edges connecting one chunk with another chunk point upwards, something that the original approach seeks to avoid, the chunks are well-separated and the overall flow of the diagram remains clearly visible. Our method is implemented in the Eclipse Layout Kernel (ELK) open-source project[1].

## 2 Wrapping Layered Graphs

A *layering* $L$ of a directed acyclic graph $G = (V, E)$ is a partitioning of the graph's nodes into distinct *layers* $L = (L_1, \ldots, L_m)$. Let $L(v)$, denote the index of $v$'s layer. In the context of Sugiyama et al.'s method, it is required that $\forall (u, v) \in E$: $L(v) - L(u) \geq 1$ holds. Furthermore, from the crossing minimization phase on a *proper* layering is required, in which all edges must be *short*: $L(v) - L(u) = 1$, $(u, v) \in E$. A proper layering can be constructed by splitting all edges that span multiple layers, introducing a *dummy node* in each spanned

---

[1] https://www.eclipse.org/elk/.

layer. A *cut index* $c$ is an index out of $[2, \ldots, m]$ where $m \geq 2$ is assumed since otherwise splitting a layering is not required. A cut index $c$ may be *valid*. The exact definition of valid depends on the use case and will be defined more precisely later on. An edge $(u, v) \in E$ is *spanning* an index $i$ with $1 < i \leq m$ if $u \in L_j, j < i$ and $v \in L_k, k \geq i$. An ordered set of monotonically increasing valid cut indices $C = (c_1, \ldots, c_n)$ partitions a layering $L$ into $n + 1$ *chunks*:

$$S_1, \ldots, S_{n+1} = (L_1, \ldots, L_{c_1-1}), (L_{c_1}, \ldots), \ldots, (\ldots, L_{c_n-1}), (L_{c_n}, \ldots, L_m).$$

Let $m' = \max_{1 \leq i \leq n+1} |S_i|$. A new layering $L' = (L'_1, \ldots, L'_{m'})$ can be constructed by subsequently adding the nodes of each $S_i$ to the layers of $L'$, each time starting at $L'_1$. An edge that spans a cut index is said to be *cut* in the new layering. To compute reasonable cut indices one must be aware of the dimensions a final drawing would have if a certain layering were to be used. During layering these dimensions can only be estimated [3]. For the remainder of this paper and for a layering $L$, let $w(L)$ and $h(L)$ denote such estimates of the layering's width and height. Now to the problem we seek to solve:

**Problem (Wrapping Layered Graphs (WLGP)).** Given a graph $G = (V, E)$, a corresponding layering $L$, and a prescribed drawing area, the problem is to find a set of valid cut indices $C$ such that for the induced layering $L'$ (1) the prescribed drawing area is used as good as possible and (2) the number of edges that point into the "wrong" direction is kept small[2].

We assess the first point using the *max scale* measure [2]. It states for a prescribed drawing area $\mathcal{R} = (r_w, r_h)$ the largest scale factor $s$ that can be used to display a graph's drawing within $\mathcal{R}$ and is computed as $s = \min \{ {}^{r_w}/_{w(L')}, {}^{r_h}/_{h(L')} \}$.

We also consider a variant of WLGP that is restricted to wrapping a single edge per cut and solely optimizes point 1), but is far easier to implement within the layer-based approach; we refer to this variant as WLGPse. Valid cuts then are in-between pairs of layers that are connected by exactly one edge: a cut index $c$ is *valid* if $|\{e \in E : e \text{ spans } c\}| \leq 1$. Next, the ideas of the heuristics we use are briefly outlined. Detailed descriptions and hints regarding the technical implementation of the overall procedure within the layer-based approach of both variants can be found in a technical report [3].

### 2.1 Heuristics

*ARD.* The *aspect ratio-driven (ARD)* heuristic selects cut indices such that the aspect ratio it estimates for the altered layering comes close to the aspect of the prescribed drawing area. The aspect ratio of the new layering $L'$ is $a = {}^{w(L')}/_{h(L')}$. The number of chunks $z$ the old layering has to be split into in order to achieve $a$ can be estimated as follows, where one assumes $w(L') \approx z \cdot w(L)$

---

[2] Note that this does not relate to the number of cuts but to the number of edges that span a certain cut index.

and $h(L') \approx {}^{h(L)}/_{z}$. Given $z$, a corresponding set of potential cut indices $C$ can be defined as well:

$$z = round\left(\sqrt{\frac{a \cdot h(L)}{w(L)}}\right), \qquad C = \left\{ \left\lceil \frac{|L|}{z+1} \cdot i \right\rceil + 1,\ \ 1 \leq i \leq z \right\}.$$

*MSD.* Aiming at a specific aspect ratio does not necessarily result in a drawing that uses a prescribed drawing area to its full potential. Hence, the *max scale-driven (MSD)* heuristic tries to directly maximize the max scale value at the cost of further work. It uses ARD's $z$ as an initial guess on a number of cut indices that would create a drawing with an aspect ratio close to the prescribed drawing area's one and then tries to find an altered layering with minimal height. The height estimate $h(L')$ we use is the maximum of sums (each sum represents the height of a single chunk, i. e. the heights of the chunk's layers). A set of cut indices that minimizes this maximum can be calculated efficiently when $z$ is given: Since the order of layers must not be altered, one can simply iteratively add the original layers to a chunk, starting a new chunk each time the current chunk's combined height exceeds ${}^{h(L)}/_{z+1}$. MSD then slightly increases and decreases the $z$ to check if a larger or a smaller number of chunks yields a better result.

*Validification and Edge-Awareness.* Remember that WLGPse, unlike WLGP, is not allowed to cut at every index, forbidden cut indices thus have to be *validified.* We evaluated two interchangeable strategies: first, incrementing a forbidden index (and all its succeeding indices) until it is valid, and second, finding the closest valid index for each forbidden index. Note that in both cases indices may have to be omitted. In the context of the unrestricted variant, "good" cuts are between pairs of layers that have a small number of edges in-between. Since neither ARD nor MSD considers this, we apply a greedy improvement strategy that tries to slightly alter the previously computed cuts in order to reduce the number of cut edges.

## 3 Results

We evaluated our method using 135 sequentialized SCGs and 146 graphs from the well-known North (AT&T) collection and found that it behaves as desired. To give an example, for prescribed drawing areas with aspect ratios between 1.0 and 4.0, MSD creates drawings that can be scaled between two and four times larger on average than when no wrapping is applied. Detailed results can be found in the technical report [3]. To conclude, we proposed an extension of the layer-based approach that creates drawings specifically tailored to prescribed drawing areas and which can easily be explained to and understood by a user. Apart from the cut points, a drawing looks similar to a traditional drawing without any cuts applied. Our method works best with sparse and "elongated" graphs and with graphs that regularly have "bottlenecks". It becomes less effective when many edges have to be cut, as is the case for dense graphs, for instance.

**Acknowledgement.** This work has been supported in part by the German Science Foundation, as part of the Compact Graph Drawing with Port Constraints (ComDraPor, DFG HA 4407/8-1) project.

## References

1. Healy, P., Nikolov, N.S.: Hierarchical drawing algorithms. In: Tamassia, R. (ed.) Handbook of Graph Drawing and Visualization, pp. 409–453. CRC Press (2013)
2. Rüegg, U., Ehlers, T., Spönemann, M., von Hanxleden, R.: Generalized layerings for arbitrary and fixed drawing areas. J. Graph Algorithms Appl. **21**(5), 823–856 (2017)
3. Rüegg, U., von Hanxleden, R.: Wrapping layered graphs. Technical report 1803, Kiel University, Department of Computer Science, February 2018. ISSN 2192–6247
4. Sugiyama, K., Tagawa, S., Toda, M.: Methods for visual understanding of hierarchical system structures. IEEE Trans. Syst. Man Cybern. **11**(2), 109–125 (1981)
5. von Hanxleden, R., Duderstadt, B., Motika, C., Smyth, S., Mendler, M., Aguado, J., Mercer, S., O'Brien, S.: SCCharts: sequentially constructive statecharts for safety-critical applications. In: Proceedings of ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2014), Edinburgh, UK. ACM, June 2014

# Diagrams, Musical Notation, and the Semiotics of Musical Composition

Vinícius Jonas de Aguiar(✉)

Centre for Philosophy of Science of the University of Lisbon,
1749-016 Lisbon, Portugal
`viniciusjonass@gmail.com`

**Abstract.** The semiotics of Charles Sanders Peirce was responsible for the blossoming of many semiotic approaches to music in the past few decades. Whilst it is clear that musicologists and philosophers of music have benefited from Peirce's semiotics to better explain and discover aspects of music, the same cannot be said about Peirce's philosophy of diagrams in particular, that seems to remain widely ignored by semioticians of music. Notwithstanding, music in general, and musical composition in particular, widely rely on schemas, rules, notational systems and other signs that might be understood as diagrams in some sense of the term. Following that clue, in this text we focus on the role played by the musical notation in the compositional process and argue that such role can be understood under the concept of diagram. We analyse an aspect of the compositional process of Beethoven's *Sehnsucht* (WoO 146) and argue that the notational system functions here as a diagram that mediates a diagrammatic reasoning process.

**Keywords:** Diagram · Musical composition · Music semiotics
Diagrammatics of music

## 1 The Concept of Diagram in Charles S. Peirce

In his well-known division of signs in 10 classes, Peirce [1, CP 2.227-264] mentions the diagram twice: to exemplify the Iconic Legisign and the Iconic Sinsign. As an Iconic Legisign, the diagram is general, a sort of law or type that is able to produce and govern replicas, i.e. the particular diagrams. The particular diagrams, in their turn, have the features of the Iconic Sinsigns. Moreover, the Interpretant of these two iconic signs is a Rheme. As Bellucci [2] puts it, the Iconic Legisign is another name for Hypoicon. It encompasses, on the one hand, the generality and structure of the Symbol and on the other, the abstraction and openness of the Icon. For instance, a particular diagram of a triangle is an Iconic Sinsign constructed based on the semiotic features of the triangles in general—the Iconic Legisign [see also 3–5].

Another source to grasp what Peirce means by diagram can be found in his famous manuscript entitled 'PAP'—more specifically in a passage quoted by Stjernfelt [3]. In this long analysis of the concept of diagram, Peirce describes it as the Interpretant of a Symbol; as such, it must encompass the generality of the Symbol as it brings its rational relations to a more sensible and particular representation through an Icon—and hence its comparison with Kant's concept of *schema*.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 748–751, 2018.
https://doi.org/10.1007/978-3-319-91376-6_73

Such a broad description obviously raises different interpretations regarding the scope of this concept, especially regarding the link between diagrams and logic, as the debate between Stjernfelt and Pietarinen shows us [6]. However, the debate just referred to might be avoided—at least for the moment—if we agree that one can also employ an extension of the concept of diagram to discuss musical semiosis. For now, we only wish to highlight the use of the musical notation as *some sort of* diagrammatic sign that allows the compositional process to be conducted in a way that very much resembles the one of diagrammatic reasoning as explained in Stjernfelt [3], as we shall indicate.

## 2 The Diagram in Beethoven's *Sehnsucht*

To avoid the complexity of the semiotic chains involved in the music composing process, we decided to focus on a rather specific part of a particular case, namely, Beethoven's sketches of the rhythmic variations for the piece *Sehnsucht* (WoO 146). It is known that Beethoven relied heavily on his sketchbooks in order to compose (for instance [7, 8]). The sketches of *Sehnsucht* are particularly attractive due to the significant number of experimentation with metrical and rhythmic patterns that were left registered by the composer.

This piece consists of vocal melody accompanied by piano, a typical *lied*. The lyrics of *Sehnsucht* are a poem by C. L. Reissig. As has been mentioned [7], by the time of this composition, Beethoven had already put into music other poems with similar verse and rhythmic structures. However, a considerable change happens in *Sehnsucht*: Beethoven uses a quite surprising time signature (3/4) which becomes central to the very melodic organisation of the vocal line. As the sketches left by Beethoven show, he tried at least *eighteen* rhythmic variations in three different time signatures (2/4; 3/4; 6/8) before deciding which one would be the right metre to compose the rhythm of the voice. Moreover, the rhythmic variations that were written to test the metric structures also appear in the final score with a few changes added. As Lockwood [7] summarises it, the notational sketches suggest that Beethoven was *testing* or *experimenting* with different metrical and rhythmic formulations composed as variations of the structure of the poem, that is, the verse pattern of 7 + 6 syllables.

Now, if we observe Beethoven's steps during the experimentation with the different time signatures, and this might be applied to the compositional process of the rhythmic structures as a whole, it is possible to notice a dynamic extremely similar, if not exactly the same, to the one described by Stjernfelt [3] as the diagrammatic reasoning process (see Fig. 1). Moreover, in the case of Beethoven's compositional process, the sign that functions as the diagram in the semiotic chain is the musical notation (see Fig. 1). We believe that music notation systems might indeed be classified as diagrams at least in a more broad sense of the term, for not only do they encompass the features of the Iconic Legisigns and Iconic Sinsigns mentioned in Sect. 1, but also their relation to their objects is visually more or less similar to the representational systems in mathematics and in logic. Musical notation systems are general and abstract signs that nonetheless have a few strict rules that guide possible transformational processes and the construction of replicas that represent their objects (music or musical ideas) solely under some relational aspects. It encompasses the representational side as well as the constructive one (as the *Schriftbildlichkeit* [9]).

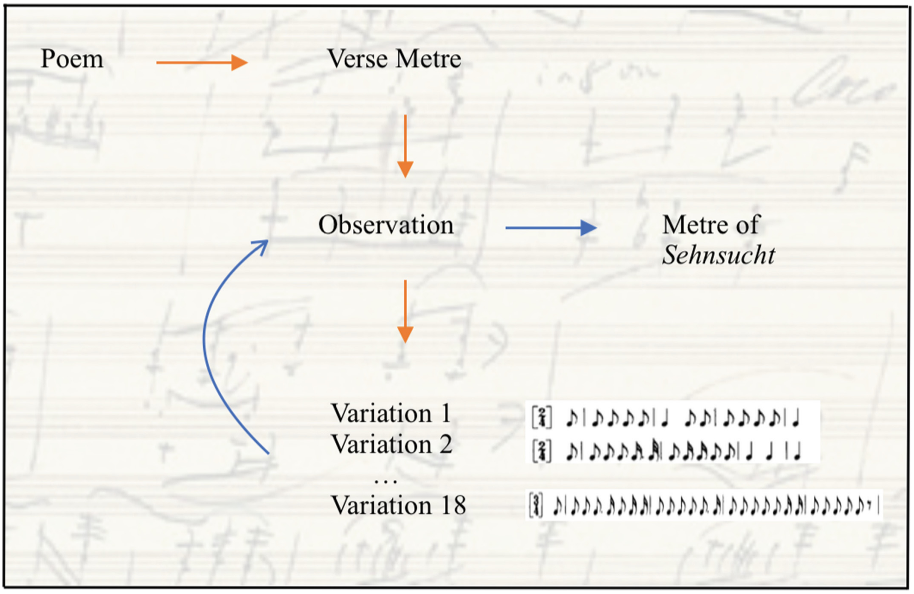


**Fig. 1.** A schema adapted from Stjernfelt [3, p. 104]. In the first step, the composer abstracts certain relations contained in the poem creating an Iconic Legisign; then, by the observation of these relations noted down in a diagrammatic way in the score (Iconic Legisign), the composer is able to manipulate the diagram and produce the variations until a definite version is reached.

The eighteen diagrammatic variations of the poem's metric pattern are just one of the many semiotic chains that led to the full composition. However, the diagrammatic semiosis, so to speak, is particularly relevant in this case. For instance, the composing of the rhythmic line sung in this *lied* consists of diagrammatic variations of the general diagram (the metrical division aforementioned). Moreover, if one considers the well-known relation between the metrical and rhythmic aspects of the *Sehnsucht* and the last movement of the *Hammerklavier* (op. 109) [7], it is clear that Beethoven transformed the metrical and rhythmic figures of *Sehnsucht* into a general diagram that was the object of several variations in the sonata *Hammerklavier*, and the same happened to certain harmonic and melodic elements (see also [10]). Taking into account the striking amount of pieces composed based on variations of other themes, patterns, rules, styles, etc., and the role of the musical notations in those cases, one can notice that the concept of diagram can be of help to approach the semiotics of musical composition.

## 3 Final Remarks

Even though very few scholars of Peirce's semiotics have noticed possible relations between music and diagrams (see for instance [11, 12]), we believe that the brief hints presented above suggest the importance of Peirce's concept of diagram to tackle the semiotics of compositional processes and possibly other aspects of music. It is also worth mentioning that due to the role of the diagrammatic sign in the experimentation and probing phases of musical composition, a taxonomy of diagrams employed in musical composition can be of great help both to composers and to music scholars in general—see, for instance, the model to compose music based on diagrams extracted from some of Kandinsky's and Mondrian's paintings [13].

## References

1. Peirce, C.: The Collected Papers of Charles Sanders Peirce, vol. 7. Harvard University Press, Massachusetts (1971)
2. Bellucci, F.: Peirce's Speculative Grammar: Logic as Semiotics. Routledge, New York (2017)
3. Stjernfelt, F.: Diagrammatology: An Investigation on the Borderlines of Phenomenology, Ontology, and Semiotics. Springer, Dordrecht (2007). https://doi.org/10.1007/978-1-4020-5652-9
4. Franco, J., Borges, P.: O Conceito de Diagrama em Peirce: uma leitura para além da gramática especulativa. Cognitio-Estudos **14**(1), 45–54 (2017)
5. Pietarinen, A.-V., Bellucci, F.: The iconic moment. towards a Peircean theory of diagrammatic imagination. In: Redmond, J., Pombo Martins, O., Nepomuceno Fernández, Á. (eds.) Epistemology, Knowledge and the Impact of Interaction. LEUS, vol. 38, pp. 463–481. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-26506-3_21
6. Pietarinen, A.-V.: Is there a general diagram concept? In: Krämer, S., Ljungberg, C. (eds.) Thinking with Diagrams: The Semiotic Bases of Human Cognition, pp. 121–138. Walter de Gruyter Inc., Boston/Berlin (2016)
7. Lockwood, L.: Beethoven: Studies in the Creative Process. Harvard University Press, Cambridge/London (1992)
8. Kinderman, W.: The Creative Process in Music: From Mozart to Kurtág. University of Illinois Press, Urbana/Chicago/Springfield (2012)
9. Krämer, S., Kirschbaum, E., Totzke, R. (eds.): Schriftbildlichkeit: Wahrnehmbarkeit, Materialität und Operativität von Notationen. Akademie Verlag GmbH, Berlin (2012)
10. Andreatta, M., Ehresmann, A., Guitart, R., Mazzola, G.: Towards a categorical theory of creativity for music, discourse, and cognition. In: Yust, J., Wild, J., Burgoyne, J.A. (eds.) MCM 2013. LNCS (LNAI), vol. 7937, pp. 19–37. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39357-0_2
11. Martinez, J.: Icons in music: a Peircean rationale. Semiotica **111**, 57–86 (1996)
12. Stweart, A.: C.S. Peirce, mechanicalism, and music. Ph.D. thesis. Texas Tech University (1987)
13. Adhitya, S., Kuuskankare, M.: Composing graphic scores and sonifying visual music with the SUM tool. In: Serafin, S. (ed.) Proceedings of the 9th Sound and Music Computing Conference, pp. 171–176. Logos Verlag, Berlin (2012)

# Calculus *CL* as Ontology Editor and Inference Engine

Jens Lemanski(✉)

KSW, FernUniversität in Hagen, 58084 Hagen, Germany
`jens.lemanski@fernuni-hagen.de`

**Abstract.** The paper outlines the advantages and limits of the so-called 'Calculus *CL*' in the field of ontology engineering and automated theorem proving. *CL* is a diagram type that combines features of tree, Euler-type, Venn-type diagrams and squares of opposition. Due to the simple taxonomical structures and intuitive rules of *CL*, it is easy to edit ontologies and to prove inferences.

**Keywords:** Diagrammatic reasoning · Knowledge Representation
Ontology Visualisation

## 1 Introduction

The idea presented here is to use Calculus *CL* as ontology editor and inference engine. *CL* was named after the so-called '*C*ubus *L*ogicus' by Johann Christian Lange, who published several similar diagrams in 1714 (cf. [1]). *CL* unifies the advantages of tree, Euler-type, Venn-type diagrams and the square of opposition. Due to its simple structure, it is especially recommended for users with little experience in ontology engineering and diagrammatic reasoning. The aim of the paper is to attract researchers for the elaboration and development of *CL*. A long-term goal is to offer *CL* as software that simplifies ontology editing and diagrammatic reasoning for non-specialists of different disciplines. In the following, only a few basic features of *CL* can be discussed.

## 2 Calculus *CL* as Ontology Editor

In their simplest taxonomical structures, good ontologies are based on the principle of JEPD, i.e. jointly exhaustive and pairwise disjoint (cf. [2]). Even the simplest structure of *CL* is based on this principle of classification and illustrates classes (A–P) and instances (Q–t) by boxes as given in Fig. 2. The taxonomical structure is for example:

$$is\_a\,(\mathrm{B},\ \mathrm{A}) \wedge is\_a\,(\mathrm{C},\ \mathrm{A}) \tag{1}$$

$$is\_a\,(\mathrm{D},\ \mathrm{B}) \wedge is\_a\,(\mathrm{E},\ \mathrm{B}) \wedge is\_a\,(\mathrm{F},\ \mathrm{C}) \wedge is\_a\,(\mathrm{G},\ \mathrm{C}) \tag{2}$$

$$\begin{gathered} is\_a\,(\mathrm{H},\ \mathrm{D}) \wedge is\_a\,(\mathrm{J},\ \mathrm{D}) \wedge is\_a\,(\mathrm{K},\ \mathrm{E}) \wedge is\_a\,(\mathrm{L},\ \mathrm{E}) \wedge is\_a\,(\mathrm{M},\ \mathrm{F}) \wedge is\_a\,(\mathrm{N},\ \mathrm{F}) \\ \wedge\, is\_a\,(\mathrm{O},\ \mathrm{F}) \wedge is\_a\,(\mathrm{P},\ \mathrm{F}) \end{gathered} \tag{3}$$

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 752–756, 2018.
https://doi.org/10.1007/978-3-319-91376-6_74

$$\textit{instance\_of}\,(\mathrm{Q},\mathrm{H}) \wedge \textit{instance\_of}\,(\mathrm{R},\mathrm{H}) \wedge \textit{instance\_of}\,(\mathrm{S},\mathrm{J}) \wedge \textit{instance\_of}\,(\mathrm{T},\mathrm{J}) \wedge \textit{instance\_of}\,(\mathrm{U},\mathrm{K}) \wedge \textit{instance\_of}\,(\mathrm{X},\mathrm{K}) \wedge \textit{instance\_of}\,(\mathrm{Y},\mathrm{L}) \wedge \textit{instance\_of}\,(\mathrm{Z},\mathrm{L}) \wedge \textit{instance\_of}\,(\mathrm{a},\mathrm{M}) \wedge \textit{instance\_of}\,(\mathrm{b},\mathrm{M}) \wedge \textit{instance\_of}\,(\mathrm{g},\mathrm{N}) \wedge \textit{instance\_of}\,(\mathrm{d},\mathrm{N}) \wedge \textit{instance\_of}\,(e,\mathrm{O}) \wedge \textit{instance\_of}\,(\mathrm{z},\mathrm{O}) \wedge \textit{instance\_of}\,(\mathrm{h},\mathrm{P}) \wedge \textit{instance\_of}\,(\mathrm{t},\mathrm{P}) \tag{4}$$

(1)–(4) can be illustrated by tree diagrams (Fig. 1) or by the cube diagram of *CL* (Fig. 2). In order to see how *CL* would look like as an ontology editor, Fig. 3 shows the famous UMLS ontology for organisms (cf. [3], slightly modified and without individuals). This would of course also work with the tree diagram in Fig. 1. But Fig. 1 needs top-down arrows to clarify the structure of the relations of classes and individuals. In contrast, *CL* (Fig. 2) uses the space of the diagram more efficiently. Thus, the arrows can be used in a different way in the diagram, e.g. to depict theorems.

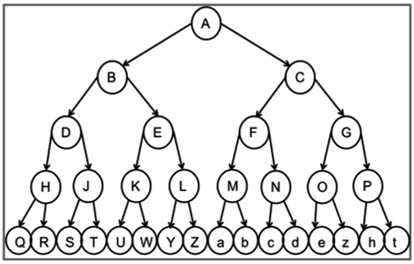


**Fig. 1.** Tree diagram

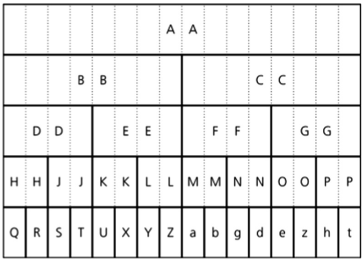


**Fig. 2.** Basic *CL* diagram

### 2.1 Arrows Depicting Theorems

In the diagram of *CL*, arrows can be used to make theorems explicit and machine-readable. The fletching of an arrow indicates the subject or the first class of a theorem and the arrowhead the predicate or the second class of a theorem. For example, the following theorems are represented in Fig. 4 by arrows (according to their color):

bottom-up arrow(s) $\quad \forall x\, \mathrm{D}x \rightarrow \mathrm{B}x$ (5)

top-down arrow(s) $\quad \exists x\, \mathrm{E}x \wedge \mathrm{K}x$ (6)

horizontal arrow $\quad \neg\exists x\, \mathrm{B}x \wedge \mathrm{C}x$ (7)

transversal arrow $\quad \exists x\, \mathrm{B}x \wedge \neg \mathrm{D}x$ (8)

(5)–(8) show that categorical propositions can be represented in *CL*. Similar to Euler-type diagrams (cf. [4], Sect. 3.2), some representations of the categorical propositions are ambiguous (e.g. cf. below Sect. 3.1). However, some of these ambiguities can also be explicitly represented in *CL*. For example, (5) implies the following Theorem (9). However, (9) can also be represented explicitly by the yellow arrows in Fig. 4:

$$\exists x \; Dx \wedge Bx \tag{9}$$

Similar to Venn-type diagrams, however, only one diagram (e.g. Fig. 2) is needed in order to draw various theorems with the help of arrows. With regard to the square of opposition, contradictory propositions are immediately prevented: For example, if Theorem (5) can be represented in *CL*, the contradictory theorem including the same classes, i.e. $\neg\exists x \, Dx \wedge Bx$, cannot be represented since it is impossible to draw a horizontal arrow (representing $\neg\exists x$) from the class D to the class B in Figs. 2 or 4.

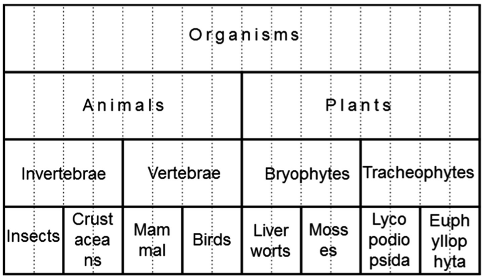


**Fig. 3.** UMLS ontology for organisms (Color figure online)

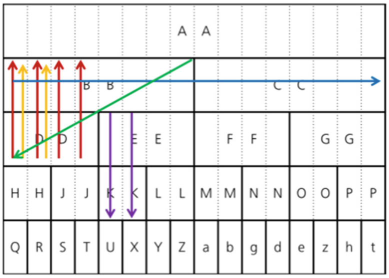


**Fig. 4.** Theorems (5)–(9) in *CL* (Color figure online)

## 2.2 Properties

Properties are best represented by a 3D visualization of *CL*. For the sake of simplicity I use a simple 2D method, in which properties are represented by a color primer in the corresponding class boxes. A diagram key outside of *CL* indicates what the color means. Properties in theorems are represented by arrows that fill only one solid box: The less colored part of the box shows the class or instance, while the more colored part shows the property. For example, Fig. 5 displays the following Theorem (10):

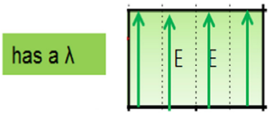


**Fig. 5.** Properties in *CL* (2D) (Color figure online)

$$\text{All E has\_a } \lambda = \forall x \; Ex \rightarrow \lambda x \tag{10}$$

# 3 Calculus *CL* as Inference Engine

*CL* can be used for various logics, e.g. predicate logic, propositional calculus or modal logic. The tasks of *CL* include **proving theorems** if the inference is complete, or **extracting information** if the inference is incomplete.

In the following, I will give just two examples of a monadic predicate calculus, both containing three theorems (i.e. three arrows) including either three classes or two classes and one property. In this case, two main rules (R) must be observed:

**(RI):** *CL* shows an inference, if two arrowheads and two arrow shafts meet in the same solid box of one class.

**(RII):** The inference of (RI) is valid, if all three theorems can be displayed in the diagram.

The first premise is given in blue, the second in green, and the conclusion in red. As example theorems I use the UMLS ontology for organisms in Fig. 3 again.

## 3.1 Example 1 (E1)

$\forall x\ \mathrm{F}x \longrightarrow \mathrm{C}x$ All bryophytes are plants. (11)

$\neg\exists x\ \mathrm{E}x \wedge \mathrm{C}x$ No vertebrate is a plant. (12)

$\neg\exists x\ \mathrm{E}x \wedge \mathrm{F}x$ No vertebrate is a bryophyte. (13)

**Theorem Proving:** As seen in Fig. 6, (11)–(13) can be represented in *CL* and (RI)–(RII) are fulfilled. Therefore, (E1) is a valid inference.

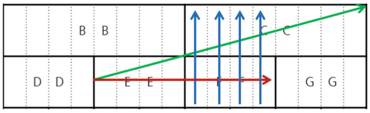


**Fig. 6.** Example 1 (E1) (Color figure online)

**Information Extraction:** Suppose Theorem (13) is unknown to the user. Since the premises, (11) and (12), are given, the software based on *CL* would have to perform forward chaining and would then have to complement the missing Theorem (13) on the basis of (RI) and (RII).

(*Ambiguity:* In Fig. 6, the transversal green arrow depicts a universal negation, i.e. $\neg\exists x$ in (12), since it is impossible to draw a bottom-up arrow from E to C.)

### 3.2 Example 2 (E2)

$\neg\exists x\ \mathrm{E}x \wedge \mathrm{D}x$ No vertebrate is an invertebrate. (14)

$\forall x\ \mathrm{E}x \rightarrow \lambda x$ All vertebrates have a vertebral column. (15)

$\neg\exists x\ \mathrm{D}x \wedge \lambda x$ No invertebrate has a vertebral column. (16)

**Theorem Proving:** Theorems (14)–(16) can be represented in *CL* (Fig. 7) and (RI) and (RII) are fulfilled. Therefore, (E2) is a valid inference

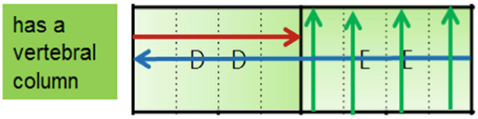


**Fig. 7.** Example 2 (E2)

**Information Extraction:** Suppose Theorem (15) is unknown to the user. Since one premise (14) and the conclusion (16) are given, the *CL*-program would have to perform backward chaining and would then have to complement the missing theorem on the basis of (RI) and (RII).

## References

1. Lemanski, J.: Logic diagrams in the weise and weigel circles. Hist. Philos. Log. **39**(1), 3–28 (2018)
2. Jansen, J.: Classifications. In: Munn, K., Smith, B. (eds.) Applied Ontology: An Introduction, pp. 159–173. Ontos, Heusenstamm (2008)
3. UMLS, https://semanticnetwork.nlm.nih.gov/. Accessed 15 10 2017
4. Bernhard, P.: Euler-Diagramme: Zur Morphologie einer Repräsentationsform in der Logik. Mentis, Paderborn (2001)

# Moded Diagrams for Moded Syllogisms

José Martín Castro-Manzano(✉) and José Roberto Pacheco-Montes

Faculty of Philosophy and Humanities, UPAEP, 21 sur 1103, Puebla, Mexico
josemartin.castro@upaep.mx

**Abstract.** In this contribution we present an extension of Englebretsen's linear diagrams in order to deal with non-classical quantifiers.

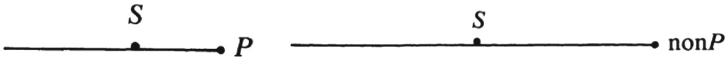
**Keywords:** Syllogistic · Non-classical quantifiers · Linear diagrams

## 1 Introduction

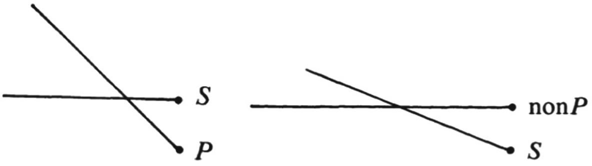
Under direct influence of [1–4], in other place we have presented an intermediate term functor logic for relational syllogistic ($\mathsf{TFL}^{+}$). Now, under direct influence of [5,6] and $\mathsf{TFL}^{+}$, we present an extension of Englebretsen's linear diagrams in order to deal with non-classical quantifiers ($\mathsf{TFL}^{\oplus}$).

## 2 $\mathsf{TFL}^{\oplus}$

Englebretsen's system requires two basic diagrammatic objects: the dot and the straight line. For instance, in order to represent the categorical propositions of syllogistic, it provides the syntax displayed in Fig. 1.



(a) All S is P (e) No S is P

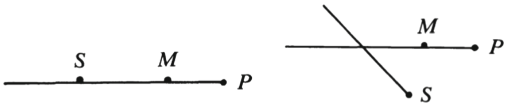


(i) Some S is P (o) Some S is not P

**Fig. 1.** Syntax for Englebretsen's linear diagrams (adapted from [5])

Given this representation, Englebretsen's system offers a clear decision procedure: a reasoning is valid if and only if the diagram of the conclusion is automatically represented after drawing down the diagram of the premises. So, for

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 757–760, 2018.
https://doi.org/10.1007/978-3-319-91376-6_75



(a) *Barbara* (b) *Darii*

**Fig. 2.** Examples of syllogisms with linear diagrams (adapted from [5])

example, valid syllogisms *Barbara* and *Darii* can be represented by the diagrams depicted in Figs. 2a and b.

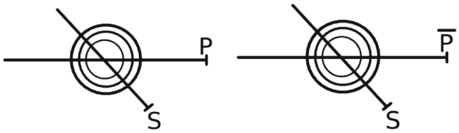
Englebretsen's diagrams provide a simple and clean diagrammatic approach for syllogistic that, alas, does not cover cases of common sense reasoning involving non-classical quantifiers such as "most," "many," or "few." Now, since we already have sentential systems to deal with this issue (for instance, $\mathsf{TFL}^+$) but we lack the corresponding diagrammatic device, we produce $\mathsf{TFL}^\oplus$, a diagrammatic system designed to capture $\mathsf{TFL}^+$. Basically, we propose a vocabulary with the next basic diagrammatic objects: the dot, the straight line, and the circle (Fig. 3). Clearly, $\mathsf{TFL}^\oplus$ is an extension of Englebretsen's diagrams.

**Fig. 3.** Vocabulary for $\mathsf{TFL}^\oplus$

With the aid of these diagrammatic objects we propose a modification of Englebretsen's linear diagrams in order to represent $\mathsf{TFL}^+$ (Fig. 4). The reason behind this modification is simple: conforming to the $\mathsf{TFL}^+$ framework, non-universal intermediate propositions are particular to some extent but have an index, and so we have to retain Englebretsen's notion of linear intersection but with the addition of circles enclosing the very intersections. This choice has the following features: since propositions a, e, i, and o have level 0 to denote the fact that they behave as usual, so the $\mathsf{TFL}^\oplus$ diagrams for a, e, i, and o have zero enclosing circles to denote the fact that they behave as usual. However, since $\mathsf{TFL}^+$ requires the new quantifiers to imply some sort of order (p (b) implies t (d), t (d) implies k (g), and k (g) implies i (o)), the superscript indexes are associated with a certain number of enclosing circles. Plus, since the indexes induce an order ($3 \geq 2 \geq 1 \geq 0$), so the number of circles induce an order ($3 \geq 2 \geq 1 \geq 0$) that indicates that a (e) does not entail p (b), t (d), k (g), i (o); but p (b), t (d), k (g) do entail i (o). Finally, since the superscript indexes are attached to both terms as to specify the detail that propositions p, t, k, b, d, and g are not convertible, so the enclosing circles are consistent with the emphasis on the scope of the terms affected by the level of quantification (so, for example, the diagram for the proposition *Few* S *are* P ($+\mathsf{S}^3 - \mathsf{P}^0$ in $\mathsf{TFL}^+$) must be read as having three circles around the intersection on S but not on P).

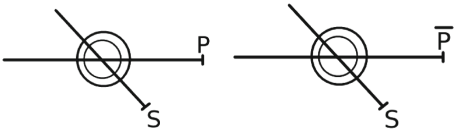
(a) All S is P $-\mathsf{S}^0 + \mathsf{P}^0$

(e) No S is P $-\mathsf{S}^0 - \mathsf{P}^0$



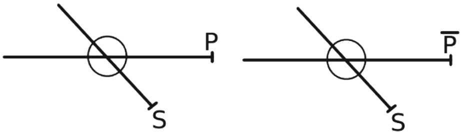
(p) Few S are not P $+\mathsf{S}^3 + \mathsf{P}^0$

(b) Few S are P $+\mathsf{S}^3 - \mathsf{P}^0$



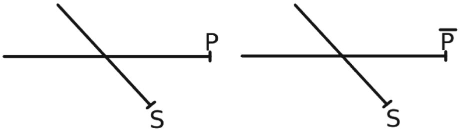
(t) Most S are P $+\mathsf{S}^2 + \mathsf{P}^0$

(d) Most S are not P $+\mathsf{S}^2 - \mathsf{P}^0$



(k) Many S are P $+\mathsf{S}^1 + \mathsf{P}^0$

(g) Many S are not P $+\mathsf{S}^1 - \mathsf{P}^0$



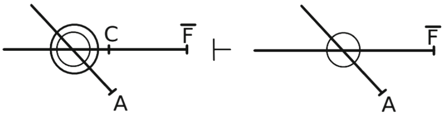
(i) Some S is P $+\mathsf{S}^0 + \mathsf{P}^0$

(o) Some S is not P $+\mathsf{S}^0 - \mathsf{P}^0$

**Fig. 4.** Syntax for $\mathsf{TFL}^{\oplus}$ and $\mathsf{TFL}^{+}$

Given this new diagrammatic representation, the adaptation of the decision procedure is trivial: a syllogism is valid if and only if the diagram of the conclusion is automatically represented after drawing down the diagram of the premises. So, for example, in the valid reasoning shown in Fig. 5, the conclusion gets clearly drawn by drawing down the premises.

As we can see, the $\mathsf{TFL}^{\oplus}$ framework gains the advantages of a diagrammatic method (a reduction of an algebraic representation to a simple and unified diagrammatic approach) and, at the same time, it gains the advantages of a theory of syllogisms with extra quantifiers (an assessment of a wide range of common sense reasoning patterns that extends the scope of traditional syllogistic). As expected, $\mathsf{TFL}^{\oplus}$ is reliable in so far as its results match those of $\mathsf{TFL}^{+}$: all valid



| Proposition | $\mathsf{TFL}^+$ |
|---|---|
| 1. No citizen is free. | $-C^0 - F^0$ |
| 2. Most adults are citizens. | $+A^2 + C^0$ |
| $\vdash$ Many adults are not free. | $+A^1 - F^0$ |

**Fig. 5.** A valid reasoning in $\mathsf{TFL}^{\oplus}$ and in $\mathsf{TFL}^+$

syllogisms in the $\mathsf{TFL}^+$ framework can be obtained by drawing down the modified diagrams of $\mathsf{TFL}^{\oplus}$, and vice versa.

## 3 Conclusions

We believe these moded diagrams are quite promising, not only as yet another set of critical thinking tools or didactic contraptions, but as a set of research devices for probabilistic reasoning (cf. [7]), psychology (cf. [8]), AI (cf. [9]), and of course, philosophy of logic (cf. [1,6,10,11]).

## References

1. Sommers, F.: The Logic of Natural Language. Clarendon Library of Logic and Philosophy. Clarendon Press/Oxford University Press, New York, Oxford (1982)
2. Thompson, B.: Syllogisms using “few”, “many”, and “most”. Notre Dame J. Form. Log. **23**(1), 75–84 (1982)
3. Peterson, P.L.: On the logic of “few”, “many”, and “most”. Notre Dame J. Form. Log. **20**(1), 155–179 (1979)
4. Mostowski, A.: On a generalization of quantifiers. Fundam. Math. **44**(2), 12–36 (1957)
5. Englebretsen, G.: Linear diagrams for syllogisms (with relationals). Notre Dame J. Form. Log. **33**(1), 37–69 (1991)
6. Englebretsen, G.: Something to Reckon with: The Logic of Terms. Canadian Electronic Library: Books Collection. University of Ottawa Press, Ottawa (1996)
7. Thompson, B.: Syllogisms with statistical quantifiers. Notre Dame J. Form. Log. **27**(1), 93–103 (1986)
8. Khemlani, S., Johnson-laird, P.N.: Theories of the syllogism: a meta-analysis. Psychol. Bull. **138**, 427–457 (2012)
9. Mozes, E.: A deductive database based on aristotelian logic. J. Symb. Comput. **7**(5), 487–507 (1989)
10. Veatch, H.B.: Intentional Logic: A Logic Based on Philosophical Realism. Archon Books, Hamden (1970)
11. Englebretsen, G., Sayward, C.: Philosophical Logic: An Introduction to Advanced Topics. Bloomsbury Academic, London (2011)

# A Review of Murner's Cards for Syllogistic

José Martín Castro-Manzano[1](✉), Jorge Medina-Delgadillo[1],
and Juan Manuel Campos Benítez[2]

[1] Faculty of Philosophy and Humanities, UPAEP, 21 sur 1103, Puebla, México
{josemartin.castro,jorge.medina}@upaep.mx
[2] Faculty of Philosophy and Letters, BUAP, Juan de Palafox y Mendoza 229-227,
Puebla, México
juancamposb@hotmail.com

**Abstract.** In this contribution we review Murner's syllogistic fragment as it appears in his *Logica memorativa*.

**Keywords:** Syllogistic · Deck of cards · Diagrams

## 1 Introduction

At the beginning of the sixteenth century we find a visual expression of Peter of Spain's *Tractatus*: Murner's *Logica memorativa Chartiludium logice siue totius dialectice memoria, & nouus Petri hyspani textus emendatus: Exercitio pictasmatis iucundo* [1]. In this contribution we focus on Murner's approach to syllogistic by describing some of his diagrams.

## 2 Murner's Cards

Between the *Tractatus* and the *Chartiludium* there are approximately three centuries of distance filled with a huge number of editions and commentaries. The commented edition of the *Tractatus* which served as the basis for Murner's *Chartiludium* was Tartaret's. Murner summarized Tartaret's explanations—matching Peter's at times—by enriching and transforming the original text which, as can be appreciated in the *Chartiludium* prologue, Murner did not find "competent." Murner summarized the major treatises and the *parva logicalia* in 51 cards (if we add the final one—which is the summary of all summaries—, we get 52 cards as in a conventional deck of cards). Each card stands for a chapter and is divided into numbered phrases. Sometimes it takes four or five phrases, sometimes more than ten, as to exhaust the doctrine of each chapter. In each card, Murner relates a sentence to a diagrammatic object (say, a gesture, an artifact, an animal) that will appear in the card along with many others elements up until the number of sentences that summarizes the chapter is reached. The diagrammatic elements

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 761–764, 2018.
https://doi.org/10.1007/978-3-319-91376-6_76

of each card are, without a doubt, striking: at times grotesque, at times enigmatic, but never dull. Some of these elements include the pole star, the tar, the rosary, the hoof of a horse, the fire, the mirrors, the pigeons, the pots, etc., and all of them stand for concepts and their positions express relation of concepts. In addition to these symbols whose purpose is to refer to the summary, there are others that give the reader the idea of which of Peter of Spain's treaties is summarized there, and given certain characters (King, Queen, Knight, Lady, ...) one can tell how far along in the *Chartiludium* one is.

## 3 Murner's Syllogistic

To pinpoint the use of Murner's cards for syllogistic we now focus on the 4th treatise by describing cards 23, 24, 25, and 26 (Fig. 1). In Fig. 1a we see one acorn. Since the acorn is the image Murner uses to represent a syllogism, this first image includes one acorn to indicate it is the first section of the fourth treatise: previous notions and remarks. In this card we see (1) the rosary beads fixed in the horse's mounth: this means the proposition, as explained in the Queen card. (2) The polar star and the wheel make an appearance once again to indicate that propositions are defined by terms, subject and predicate terms. (3) and (4) represent the sun that relates universal quantification to either term.

In Fig. 1b we see two acorns: we are in the second section, the one that indicates the general structure of a syllogism. We see (1) three cats, because a syllogism requires three categorical propositions. Then we see (2) a spear that is greater to the left (2a) to indicate the major term, has a middle section to indicate the middle term (2c), and is pointy to the right (2b) to indicate the minor term. (3) indicates the three propositions. (4) is a mirror that stands for the figure of the syllogism. (5) is the balance that represents the moods. (6) is the sun that captures the notion of quantity, while (7), the stone, represents quality. Finally, (7) the five feathers above the cap represent the five rules of validity.

In Fig. 1c we see three acorns: we are in the third section, the one that indicates the structure of the syllogisms of the first figure. Notice this card has no arabic numbers: we can see the mirror that denotes the figure, but now with a roman number, *I*, to indicate this card represents the syllogisms of the first figure. The key indicates that the conclusion follows directly (i.e., the minor term is the subject and the major term is the predicate) whereas the lock pick (hidden in the gentleman's back) indicates that the conclusion follows indirectly. The feathers above the cap indicate we require two rules of validity for the syllogisms of the first figure.

In Fig. 1d we see four acorns: we are in the fourth section, the one that indicates the structure of the syllogisms of second and third figures. Notice this card has no arabic numbers as well: we can see a mirror to the left (in the lady's right hand) with the roman number *III* that denotes the third figure of the syllogism, while to the right (in the lady's left hand) we see a mirror with the roman number *II*, which denotes the second figure of the syllogism. Notice also that, again to the left there are two feathers while to the right there are also two feathers, indicating the corresponding rules of validity. The table hanging in the lady's chest indicate the moods. And the four bees above the lady are going back to their queen as to indicate that these figure can be reduced to the syllogisms of the first figure, as in *Pr. An.* A.1, 25b1.

(a) Card 23. Previous notions (b) Card 24. General structure

(c) Card 25. First figure (d) Card 26. Second and third figures

**Fig. 1.** Murner's summary of syllogistic

## 4 Conclusions

That Murner's cards for syllogistic are logic diagrams [2, p. 28] may be called into question, but it is safe to say, after this quick review, that we have enough elements to claim that Murner's cards for syllogistic do constitute a system of diagrams *stricto sensu* nonetheless. Consider that *(i)* they convey spatial information just as *bona fide* diagrams do: indeed, the cards contain all the basic tenets of syllogistic stored at one particular *locus* by also including information about relations with the adjacent *loci* [3]; and so *(ii)* they summarize sentential information just as diagrams do: each card groups together a lot of information thus avoiding large amounts of search [3]. *(iii)* The cards work as instruments [4] that promote system understanding just as any set of legitimate diagrams would do in so far as they offer a visual explanation of the structure of syllogistic. This visual explanation is clear: by using compositionality and having a clear motivation, *(iv)* Murner's cards guarantee the diagrammatic objects are not semantically overloaded and they allow a hierarchical description of syllogistic [5]. *(v)* Finally, Murner's cards support the process of explaining or teaching the general structure of a sentential system, just as typical diagrams do. For these reasons, we think it is fair to say that Murner's system deserves not only a place in the history of logic, but in the history of diagrams.

## References

1. Murner, T.: Logica Memorativa. The Bavarian State Library, München (1509)
2. Gardner, M.: Logic Machines and Diagrams. University of Chicago Press, Chicago (1982)
3. Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. Cognit. Sci. **11**(1), 65–100 (1987)
4. Moktefi, A.: Diagrams as Scientific Instruments, pp. 81–89. Peter Lang Verlag, Frankfurt (2017)
5. Nakatsu, R.T.: Diagrammatic Reasoning in AI. Wiley, Hoboken (2009)

# Towards Executable Representations of Social Machines

Dave Murray-Rust[1](✉), Alan Davoust[1], Petros Papapanagiotou[1], Areti Manataki[1], Max Van Kleek[2], Nigel Shadbolt[2], and Dave Robertson[1]

[1] University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, UK
{d.murray-rust,adavoust,pe.p,a.manataki,dr}@inf.ed.ac.uk
[2] University of Oxford, Parks Rd, Oxford OX1 3QD, UK
{max.van.kleek,nigel.shadbolt}@cs.ox.ac.uk

**Abstract.** Human interaction is increasingly mediated through technological systems, resulting in the emergence of a new class of socio-technical systems, often called *Social Machines*. However, many systems are designed and managed in a centralised way, limiting the participants' autonomy and ability to shape the systems they are part of.

In this paper we are concerned with creating a graphical formalism that allows novice users to simply draw the patterns of interaction that they desire, and have computational infrastructure assemble around the diagram. Our work includes a series of participatory design workshops, that help to understand the levels and types of abstraction that the general public are comfortable with when designing socio-technical systems. These design studies lead to a novel formalism that allows us to compose rich interaction protocols into functioning, executable architecture. We demonstrate this by translating one of the designs produced by workshop participants into an a running agent institution using the Lightweight Social Calculus (LSC).

**Keywords:** Social machines · Diagrammatic interface
Rapid assembly · Prototyping

## 1 Introduction

Ubiquitous computation and digital communication systems have produced new forms of socio-technical systems, vast networks where people achieve coordinated action at scale. Examples include Wikipedia, Twitter, Ushahidi and so on. Characterising such systems requires thinking beyond their software infrastructure: one unified lens for viewing them is Berners-Lee's concept of *social machines* [2,7], that describes systems in which humans and computation play complementary roles.

In this paper, we explore the design and construction of social machines through a diagrammatic language. Our main design goals for the language are (i) to be accessible to non-specialists, enabling them to craft their own social

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 765–769, 2018.
https://doi.org/10.1007/978-3-319-91376-6_77

machines themselves, and (ii) to be sufficiently formal to be able to turn into executable code, so that designers can create running systems directly from diagrams. The language we propose has been shaped by a series of design workshops where non-expert participants designed Social Machines with limited guidance, using some suggested constructs based on the notions of roles and protocols – and their own free-form diagramming skills.

From the outcomes of these workshops we designed a diagrammatic language and methodology to design Social Machines as Electronic Institutions, formalized with the Lightweight Social Calculus (LSC [5]) to provide executable infrastructures. We present an intuitive design process for users to construct Social Machine models using this formalism, and demonstrate the execution of a model using a generic LSC engine.

## 2 Design Method

Our language has been refined based on four participatory design workshops, mostly involving non-experts (who were new to the concept of social machines), where participants collaboratively designed their own social machines from scratch, using paper and markers. At the beginning of each session, participants were provided handouts with example diagrammatic primitives that we designed (Fig. 1a), that they could (optionally) use to help them sketch their social machines.

Figure 1b shows one ad-hoc diagram created by non-specialists, showing roles, coordination, interactions, implementation hints and social aspects of the system concisely and comprehensibly. From such diagrams we extracted the following key elements: *People*, playing a range of roles: restaurants that provide waste food, informants who spot people in need, *madres* that mediate the interactions and so on; *Infrastructure*, whether physical or computational, that coordinates the activity of humans around their purpose; and finally *connecting arrows*, representing the transmission of messages or physical objects (‘gives food’); more general geo-spatial interactions (‘detects’); and bundles of possible operations on computational systems (‘make announcements’) that implicitly include access control, posting, retracting and updating information etc.

At the conclusion of each session, each group was asked to present their social machine; photos were captured of diagrams made, and verbal descriptions were recorded, transcribed, and archived. We then analysed the graphical vocabulary used in the diagrams, along with descriptions, to identify where graphical primitives were appropriated, reused, and extended. This enabled us to identify robust components, eliminate and refine components that were not used or misunderstood. For the final workshop, we presented the participants with a paper based version of the diagram tool discussed in this paper, prompting them to investigate and define the interactions between actors.

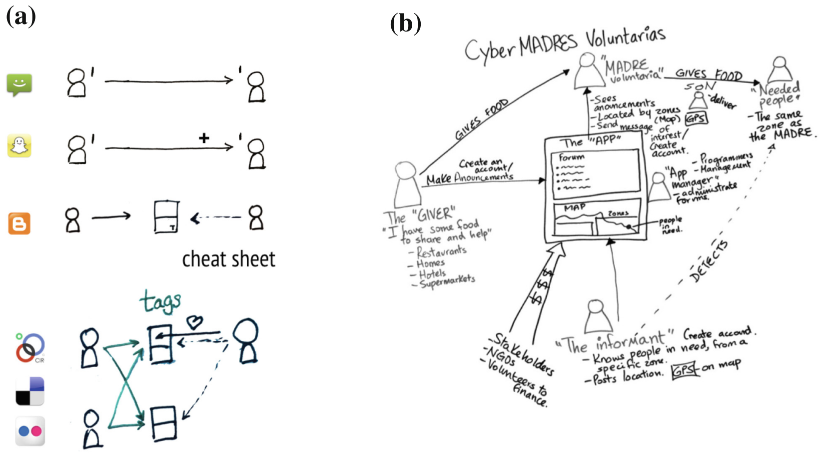


**Fig. 1.** (a) Some of the graphical elements from the first Sociograms workshop. (b) The *Cybermadres* social machine, sketched by workshop participants. This Social Machine would support the activities of volunteers in Mexico, who collect excess food from restaurants and distribute it to people in need.

## 3 Diagram Language

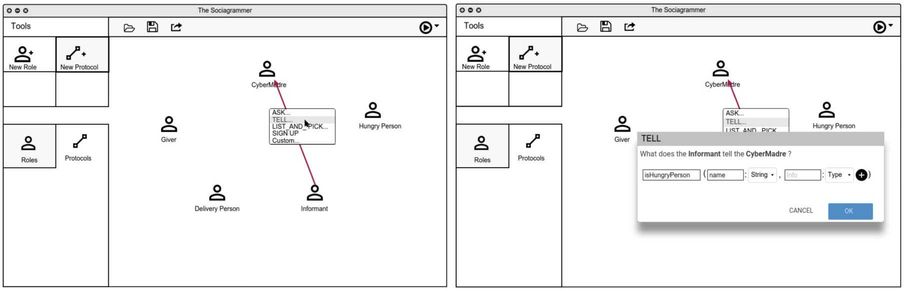
Our diagrammatic language contains the following elements, with a full demonstration given in Fig. 2:

**Nodes** represent actors in the system, whether computational or human.
**Edges** define the interactions between these actors, by specifying interaction protocols (Fig. 2a).
**Protocols** are specified as generic activities (e.g. ASK, ESCROW), and specialised to define the kinds of data that flows through them. For example, a simple communication protocol might carry a specification of the kinds of messages to be transferred, which then implies that the roles involved in the interaction are capable of providing or processing that type of information.

The diagram shown in Fig. 2c details the key interactions of a particular social machine infrastructure—in this case, a system that supports the *CyberMadres* example in Fig. 1b. Our interface mock-ups sketch the design process and the extra information needed to create a working system.

### 3.1 From Diagrams to Executable Systems

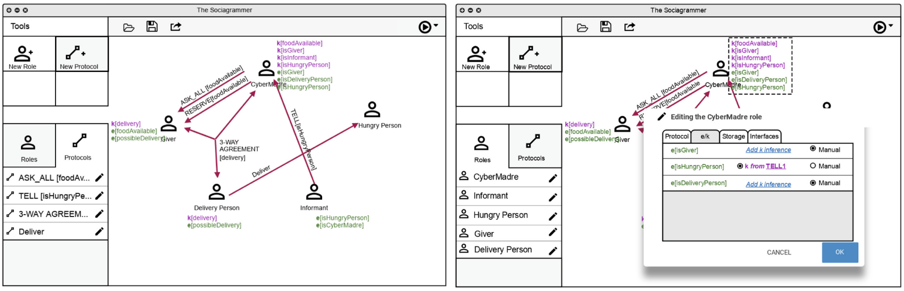
Diagrams created in this manner can be automatically transformed into the Lightweight Social Calculus (LSC) [4,6], a high level protocol language. An LSC protocol consists of the roles that each participating agent may play, by executing the part of the protocol that corresponds to their role locally. Each role may involve message passing and computation, tied together with conditions (`If` ... `then`) and temporal sequencing (`Then` and `Or`). Computation involves input



What does the **Informant** tell the **CyberMadre** ?

(a) Definition of actors in the system, with protocols describing interactions between them

(b) Specialisations of the protocols to carry appropriate data for the given interaction



(c) A complete system, with multi-way interactions and composed capabilities

(d) Eliciting connections between the inputs and outputs of different actors

**Fig. 2.** Example operation of the Sociogrammer tool, from initial specification of actors through to linking predicates

predicates `e()`, and output predicates `k()`. Inputs *elicit* information, typically through a local knowledge base, or by "asking a human" via an interface, at which point human decision making can enter the computational system. Output predicates indicate that the agent now *knows* something, which implies that the agent be capable of processing the information, e.g. storing it in a local knowledge base or otherwise changing the state of the world in a representative way.

The design process (Fig. 2d) involves connecting inputs from one protocol to outputs of another, or else flagging particular inputs to be fulfilled by processes outside the scope of the system.

## 4 Discussion and Conclusions

There are two key questions behind this work: (i) Can we develop a diagrammatic language for designing social machines, accessible to non-experts? and (ii) Can we design this language in such a way that it produces executable infrastructures?

A key design challenge is to find the correct, composable units to build these diagrams from. Here we have used interaction protocols, as a way to cover complex patterns of activity that unfold in time using simple, human identifiers, e.g. “arrange a meeting”. Our workshops have shown that participants see real possibilities in designing such systems, and their ad-hoc diagrams show some of the key concepts needed to describe interaction protocols. Participants to produce meaningful and plausible social machine designs in a 2–3 h workshop, from a standing start. Designs included social interventions that help the homeless, shared diaries for nomads, crowdsourced traffic reports and interpersonal archives.

Relative to the second question, we have shown a prototype interface, where a version of the simple, at-hand iconography can be used to specify enough detail to create executable infrastructures. This demonstrates the potential for a translation from simple readable diagrams into working systems.

The question remains as to what a meaningful set of interaction primitives might be, simple yet expressive enough to describe most social machines. Our initial analysis has brought up a range of useful constructs, which allowed us to fulfil the designs created by workshop participants.

This leaves us with a form of extremely concise Model Driven Development [1,3] that supports a democratic, participatory approach that allows a wide range of people to design a profusion of small social machines, adapted to their particular communities of practice.

This proof of concept indicates that there is a level of representational complexity that allows people to make their intentions clear about complex systems that would otherwise be beyond their ability to design.

## References

1. Agüero, J., Rebollo, M., Carrascosa, C., Julián, V.: MDD for virtual organization design. In: Demazeau, Y., et al. (eds.) Trends in Practical Applications of Agents and Multiagent Systems. Advances in Intelligent and Soft Computing, vol. 71, pp. 9–17. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12433-4_2
2. Berners-Lee, T., Fischetti, M., Foreword By-Dertouzos, M.L.: Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor. HarperInformation, New York City (2000)
3. France, R., Rumpe, B.: Model-driven development of complex software: a research roadmap. In: 2007 Future of Software Engineering, pp. 37–54. IEEE Computer Society (2007)
4. Murray-Rust, D., Papapanagiotou, P., Robertson, D.: Softening electronic institutions to support natural interaction. Hum. Comput. **2**(2), 34 (2015)
5. Murray-Rust, D., Robertson, D.: LSCitter: building social machines by augmenting existing social networks with interaction models. In: Chung, C., Broder, A.Z., Shim, K., Suel, T. (eds.) 23rd International World Wide Web Conference, WWW 2014, Seoul, Republic of Korea, 7–11 April 2014, Companion Volume, pp. 875–880. ACM (2014). http://doi.acm.org/10.1145/2567948.2578832
6. Robertson, D.: A lightweight coordination calculus for agent systems. In: Leite, J., Omicini, A., Torroni, P., Yolum, I. (eds.) DALT 2004. LNCS (LNAI), vol. 3476, pp. 183–197. Springer, Heidelberg (2005). https://doi.org/10.1007/11493402_11
7. Shadbolt, N., Van Kleek, M., Binns, R.: The rise of social machines: the development of a human/digital ecosystem. IEEE Consum. Electron. Mag. **5**(2), 106–111 (2016)

# Transforming Storyboards into Diagrammatic Models

Elena-Teodora Miron(✉), Christian Muck, Dimitris Karagiannis, and David Götzinger

Research Group Knowledge Engineering, University of Vienna,
Währinger Str. 29, 1090 Vienna, Austria
`{elena-teodora.miron,christian.muck,`
`dimitris.karagiannis,david.goetzinger}@univie.ac.at`

**Abstract.** Design Thinking can be employed to define services, new product (features), innovative processes and disruptive business models collaboratively for digitization. Diagrammatic models play an essential role here as they capture relationships between different aspects of the problem. When computed by means of software they also explicitly show details which in design thinking tools users implicitly fill with their own world-understanding, thus fostering a clear and transparent representation of the problem space. In addition, diagrammatic models can be enriched by semantics and subsequently be queried, analysed and processed.

The paper at hand shows the DigiTrans (http://www.interreg-danube.eu/approved-projects/digitrans) project approach for an automated transformation process of haptic storyboards into diagrammatic models by means of video-imaging and web-services.

**Keywords:** Design Thinking · Storyboards · Diagrammatic models

## 1 Introduction

Digitization greatly challenges small and medium enterprises (SMEs) to innovate their products/services and/or their business model to remain competitive. Yet strapped for resources they have difficulties in facing these challenges. DigiTrans, an EU-funded project, aims to offer a mixed approach of training and incubation, to increase innovation capabilities of SMEs and support them through the digital transformation process.

Because the problems faced in the digital transformation process tend to be complex and vaguely defined the project recommends different Design Thinking based tools to visualize the problems and to generate innovative solutions. Some of these tools focus on storyboarding as a means to aid innovation. Storyboards are a sequence of scenes that represent the main points of a story. They are designed to communicate intuitively high-level ideas to the viewer. Yet how different scenes relate together and what the exact semantic meaning and importance of each object in a scene is remains largely at the understanding of each individual visualizing it. As shown by [3] in such cases viewers tend to implicitly fill the missing information with their own

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 770–773, 2018.
https://doi.org/10.1007/978-3-319-91376-6_78

world-understanding. Yet when aiming to materialize an innovative idea presented in a storyboard different aspects of the problem and the relations between them need to be unambiguously described in order to avoid conflicts. Transforming the haptic storyboard by means of an automated import into diagrammatic models and facilitating semantic enrichment (e.g. by describing object properties) can help increasing information transparency, clarity for model users and model consistency.

The remainder of this paper describes a real-world digital transformation scenario (Sect. 2), the environment provided for prototyping (Sect. 3) and the application experience (Sect. 4).

## 2 The Digital Transformation Scenario

A manufacturer (an SME) of cleaning products is facing a steep revenue decline. An initial analysis (a series of ideation workshops) has produced the following insights: the company's strong points are its products and customer knowledge. The weak point is the distribution via direct marketing, which the company uses since the 1970's, and which nowadays represents a problem due to the changing structure of households.

The company has identified the opportunity to use online distribution via a mobile app/website to attract a new group of customers. They defined the key customer via a persona description: young women (between 25 and 45), family or double-income-no-kids, disposable income, environmentally aware, interested in the quality of products and low-pollutants ingredients, working. Such a female must be able to choose between a catalogue and electronic order. In the electronic channel she can select the delivery date, place and time according to her availability in addition to her products.

As a next step a storyboard needs to be defined for the interaction between the customer and the company's online channel/mobile app. These scenes represent subsequently interaction points for the company's processes and IT services.

## 3 Developing the Scene2Model Service Environment

Scene2Model is an end to end process for the software-supported transformation from tangible figures to diagrammatic models with the possibility of simultaneous semantic enrichment of objects. This and the automatic composition of models into storyboards innovate the way diagrams can be used in industry work environments.

Several preparation and development steps were necessary before the Scene2Model service could be used in a workshop setting for design thinking prototypes.

***1. Prepare the incubation space*** – this includes a table with a transparent top, a web-enabled camera underneath it and haptic storyboard figures with tags glued to their bottoms.

During workshops haptic figures will be positioned on a table. They depict the scenes of a storyboard. These haptic figures are intuitively understandable, easily rearrangeable and help to facilitate the mental design process of people, by allowing them to interact with a physical representation of their thoughts [5]. In our prototypical implementation we used SAP Scenes [4] storyboard figures.

The tags, i.e. unique IDs, glued at the bottom of each figure are used to identify and map the haptic objects to modelling objects (of a modelling application) as well as to calculate their relative position on the table. The camera positioned underneath the table records a live video stream of the tags. The ID and the coordinates are offered over a network interface, which can be used by other applications.

***2. Develop a modelling method and an ontology for transforming the haptic scenes into diagrammatic models***

The prerequisite for the automatic transformation was the development of a modelling method [2] describing the haptic figures. Modelling methods can facilitate the Design Thinking process (cf. [1]). For the prototype a meta-model was created to define syntax, notation and semantics of the different haptic figures or objects. Each meta-model class possesses multiple graphical presentations, which are defined by the *Type* of an object. *Type* is a property, just like the properties *Name* and *Description*. The first one defines an identifier for a modelling object and the second one holds a natural language description. Classes with a *Text* property, show this text directly in diagrammatic model. The class *Character* also implements the properties *Role* and *Age*, which can be used to specify the humans in the modelled scenario.

An ontology is used to map the IDs, from the tag recognition to modelling objects. Therefore, the information from the meta model is saved in the ontology and enhanced with a *TagID* property.

The described meta model was implemented in the ADOxx[1] metamodeling platform, which supports the definition of classes and their properties, the implementation of mechanisms as well as the use of a built-in modelling toolkit, where defined models can be used.

A mechanism implementation is used to trigger the Scene2Model software directly from the ADOxx modeling tool. Data is gathered from the tag recognition software. With the received ID, the additional information is read from the ontology. Then the combination of the ontological information and the position (from the tag recognition) is sent back to ADOxx, where the modelling objects are created and positioned automatically.

## 4 Applying the Scene2Model Service

The first step in the workshop is the creation of the storyboard. Participants identified ordering via app using a tablet at home as a key scene. The different objects were arranged accordingly. The description of the customer interaction with the app was pinned on handwritten notes on a meta-plan board. As soon as the participants agreed on the scene design, ADO*xx* was triggered and a computable diagrammatic model was created automatically. The results of both steps are shown below in Fig. 1.

Subsequently participants reviewed the model and entered different properties and explanatory descriptions to the different objects to make the scene more descriptive. Generating and updating models from the new storyboard layout is possible at any

---

[1] https://www.adoxx.org/live/home.

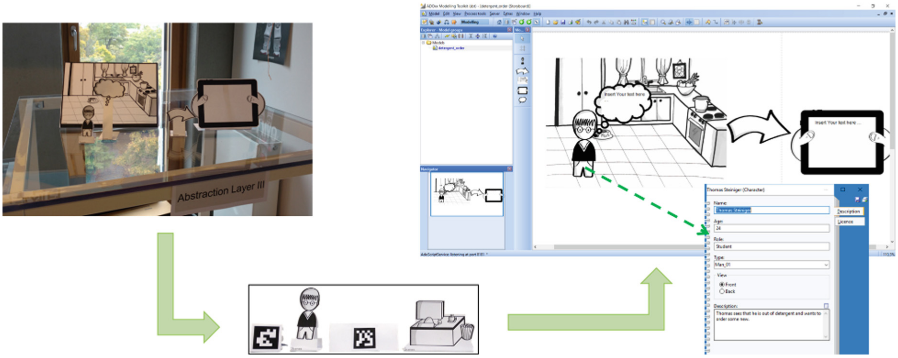


**Fig. 1.** Scene2Model service application

time. The differences, at class and attribute level, between the original and the modified storyboard can be queried by mechanisms provided out of the box by the platform.

The workshop participants were positively receptive towards the perceived benefits of the solution, especially in lieu of the time and location-independent preservation of the workshop results as well as their usability to further enhance scenarios by introducing information into the diagrammatic models directly and collecting all information in one place. Further evaluations of the usability and value of the implementation will be done during 2018 in different Central European countries with SMEs.

**Acknowledgement.** This paper and the work contained therein were funded by the Danube Transnational Program through the DigiTrans (DTP1-193-1.2) project.

## References

1. Bork, D., Karagiannis, D., Hawryszkiewycz, I.T.: Supporting customized design thinking using a metamodel-based approach. In: 20th Australasian Conference on Information Systems (ACIS) 2017, pp. 1–11, December 2017. http://eprints.cs.univie.ac.at/5314/
2. Karagiannis, D., Kühn, H.: Metamodelling platforms. In: Bauknecht, K., Tjoa, A.M., Quirchmayr, G. (eds.) EC-Web 2002. LNCS, vol. 2455, p. 182. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45705-4_19
3. Pérez-Kriz, S., Vivancos-Pérez, R.F.: Using diagrammatic drawings to understand fictional spaces: exploring the buendía house in Gabriel García Márquez's *One Hundred Years of Solitude*. In: Jamnik, M., Uesaka, Y., Elzer Schwartz, S. (eds.) Diagrams 2016. LNCS (LNAI), vol. 9781, pp. 187–193. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42333-3_14
4. SAP SE: Scenes. https://experience.sap.com/designservices/approach/scenes. Accessed 7 Dec 2017
5. Tschimmel, K.: Design thinking as an effective toolkit for innovation. In: ISPIM Conference Proceedings, p. 1. The International Society for Professional Innovation Management (ISPIM) (2012)

# Logical Reasoning with Object Diagrams in a UML and OCL Tool

Khanh-Hoang Doan(✉) and Martin Gogolla

Computer Science Department, University of Bremen, 28359 Bremen, Germany
{doankh,gogolla}@informatik.uni-bremen.de

**Abstract.** In this contribution, we introduce an approach to visualize and analyze logical reasoning problems in a UML and OCL tool by using logical puzzles represented with UML diagrams. Logical reasoning is formalized as a UML class diagram model enhanced by OCL restrictions. Puzzle rules and questions are expressed as either partial object diagrams or OCL formulas within the model. Solutions can be found and explored by a tool as object diagrams.

**Keywords:** Logical reasoning · UML and OCL model
Object diagram

## 1 Introduction

UML diagrams, such as class and object diagrams, are utilized to diagrammatically represent real-world system at an abstract level with some constraints formulated in OCL. Taking this as a basis, UML and OCL can be a promising solution for representing and visualizing logical reasoning problems. One approach for that will be introduced in this paper. A logical reasoning problem is represented as a UML class model enhanced by OCL restrictions. Rules and questions are formulated as either partial diagrams or OCL formulas within the model. The solutions can be found using a deduction system integrated in a tool and represented as object diagrams. This contribution focuses on representation and analysis of logical reasoning problems, in the context of the tool USE (Uml-based Specification Environment) [2].

Recently, the application of diagrammatic representation for reasoning has been a widely considered topic. For example, the approaches for reasoning with diagrams, i.e., Euler, Spider diagrams and Graphs, have been presented recently in [5], [6] and [4], respectively. In contrast to these approaches, our contribution uses UML diagrams, i.e. class and object diagrams, to visualize and analyze logical reasoning problems. In next section we will illustrate the idea of representing and visualizing a logical reasoning problem with our tool USE. In the last section, the paper ends with concluding remarks.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 774–778, 2018.
https://doi.org/10.1007/978-3-319-91376-6_79

## 2 Solving and Representing Logical Reasoning Problems

**Running Example:** In order to demonstrate our approach, an example of a logical reasoning problem is discussed. The deduction problem that we introduce here is 'Einstein's Puzzle', a very well-known logic puzzle which sometimes is used as an example in teaching logic and formal methods [3].

**Fig. 1.** Einstein's puzzle.

The problem can be described as follows: Let us assume that there are five houses of different colors next to each other on the same road. In each house lives a man of a different nationality. Every man has his favorite drink, his favorite brand of cigarettes, and keeps a particular pet. There are fifteen clues of deduction that are listed below, and Fig. 1 illustrates the reasoning problem.

01. The Briton lives in the red house.
02. The Swede keeps dogs as pets.
03. The Dane drinks tea.
04. Looking from in front, the green house is just to the left of the white house.
05. The green house's owner drinks coffee.
06. The person who smokes Pall Malls raises birds.
07. The owner of the yellow house smokes Dunhill.
08. The man living in the center house drinks milk.
09. The Norwegian lives in the leftmost house.
10. The man who smokes Blends lives next to the one who keeps cats.
11. The man who keeps a horse lives next to the man who smokes Dunhill.
12. The owner who smokes Bluemasters also drinks beer.
13. The German smokes Prince.
14. The Norwegian lives next to the blue house.
15. The man who smokes Blends has a neighbor who drinks water

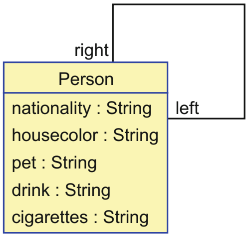


**Fig. 2.** Model of the logical reasoning problem.

**Construct a Model:** Firstly, a model corresponding to the problem being solved must be constructed. Depending on the problem, the model must include all necessary information, i.e., classes, attributes and associations, so that the model can simulate the problem.

For Einstein's Puzzle, we have formulate a model with five attributes: nationality, housecolour, pet, drink, cigarettes, and one association that represents the neighborhood relationship between the persons. Figure 2 presents the class diagram of the model.

**Formulate Invariants:** After constructing a suitable model, a collection of OCL invariants must be formulated on the model, one invariant for each clue (rule). As mentioned before, it is important that the model must cover all information from all clues. Therefore, we formulate 15 invariants corresponding to 15 rules of the puzzle. For instance, the following listing is the invariant corresponding to the rule 08 “The man living in the center house drinks milk”.

```
context Person inv clue08:
  Person.allInstances()→one(p|Set{p} →closure(left)→size()=Set{p}
  →closure(right)→size() and p.drink='Milk')
```

In addition to the textual representation, some rules can be represented as a partial object diagram, which can enhance the understandability. For example, the diagrammatic visualization for the rule 08 is shown in Fig. 3.

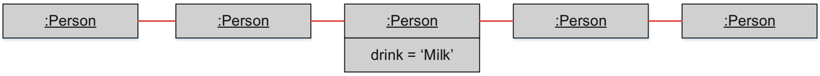


**Fig. 3.** The diagrammatic visualization of rule 08.

The other invariants are formulated analogously. The full model including all invariants is presented in [1].

**Solving the Problem with the Model Validator:** After constructing a suitable model with all necessary invariants corresponding to all clues, we apply the model validator to solve the problem and find the answer. In the case of Einstein’s puzzle, the model validator finds a solution as an object diagram, which is shown in Fig. 4. We arranged the 5 persons (objects) according to their neighborhood relationships from left to right. In the found solution we can easily check that some simple rules, e.g., rules 01, 02, 03, 13, are satisfied. Further analysis of the satisfaction of more complicated rules (invariants) on the found solution can be done with our tool with the ‘Evaluation browser’ functionality.

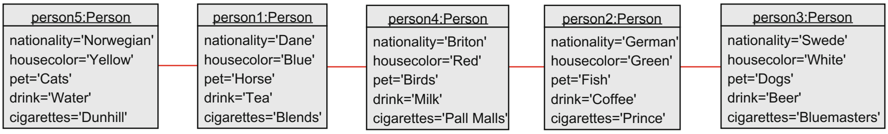


**Fig. 4.** Found solution for Einstein’s puzzle.

The search space of the model validator is defined by a configuration. Therefore we can speed up the solving process by setting a suitable configuration. Setting a proper configuration plays an essential role in the context of solving reasoning problem with the model validator. To archive this, one can go through the description of the problem and underline the number of objects (man/person) and the values which are given corresponding to the class attributes. As a result, the following list is the configuration for Einstein’s puzzle.

```
Person_min_max = 5..5
Person_nationality =
  Set{'Norwegian','Dane','Briton','German','Swede'}
Person_housecolor = Set{'Red','Yellow','Blue','Green','White'}
Person_pet = Set{'Cats','Birds','Horse','Fish','Dogs'}
Person_drink = Set{'Water','Tea','Milk','Coffee','Beer'}
Person_cigarettes =
  Set{'Dunhill','Prince','Blends','Pall Malls','Bluemasters'}
Connecting_min_max = 4..4
```

**Explore Solution Universe:** In a case of having more than one solution, the model validator also provides an option to explore all of them. Naturally, this will be possible only if the solution universe is relatively small. To achieve all solutions we use the command `mv -scrollingAll <PropertyFile>`, and the additional succeeding command `mv -scrollingAll [prev|next|show(<N>)]` allows us to scroll through the solution interval and show each of them as an object diagram. For example, after executing the command `mv -scrollingAll show(1)`, the first (and only) solution is shown as an object diagram presented in Fig. 4.

## 3 Conclusion and Future Work

In this contribution we have described our method for visualizing and analyzing logical reasoning problems in the tool USE. We have used diagrammatic representations and puzzles as a cheap-prized entry to formal methods. Beside the Einstein's Puzzle, we have been applied the introduced approach for several popular logical reasoning examples, e.g. Sudoku puzzles or scheduling problems; the details can be seen in [1]. As future work we have identified to handle the various puzzle examples present in the literature in our approach. A further option is to develop a particular USE plugin particularly aiming at puzzle representation and to handle their solutions, as well as to allow the specification of OCL expressions with (partial object) diagrams.

## References

1. Doan, K.H., Gogolla, M.: Addendum to logical reasoning with object diagrams in a UML and OCL tool. Technical report, University of Bremen (2017). http://www.db.informatik.uni-bremen.de/publications/intern/ReasoningwithUSE.pdf
2. Gogolla, M., Hilken, F.: Model validation and verification options in a contemporary UML and OCL analysis tool. In: Proceedings of Modellierung (MODELLIERUNG 2016). GI, LNI, vol. 254, pp. 203–218 (2016)
3. Spichkova, M.: "Boring formal methods" or "Sherlock Holmes deduction methods"? In: Milazzo, P., Varró, D., Wimmer, M. (eds.) STAF 2016. LNCS, vol. 9946, pp. 242–252. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-50230-4_18
4. Takemura, R.: A logical investigation of heterogeneous reasoning with graphs in elementary economics. In: Jamnik, M., Uesaka, Y., Elzer Schwartz, S. (eds.) Diagrams 2016. LNCS (LNAI), vol. 9781, pp. 98–104. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42333-3_8

5. Urbas, M., Jamnik, M., Stapleton, G., Flower, J.: Speedith: a diagrammatic reasoner for spider diagrams. In: Cox, P., Plimmer, B., Rodgers, P. (eds.) Diagrams 2012. LNCS (LNAI), vol. 7352, pp. 163–177. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31223-6_19
6. Zelazek, F.: Diagrammatically explaining Peircean abduction. In: Dwyer, T., Purchase, H., Delaney, A. (eds.) Diagrams 2014. LNCS (LNAI), vol. 8578, pp. 308–310. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44043-8_30

# Scanning the Invisible: Framing Diagrammatic Cognition in Experimental Particle Physics

Javier Anta(✉)

LOGOS - Universitat de Barcelona, Barcelona, Spain
antajav@gmail.com

**Abstract.** In this study I aim to develop a cognitive evaluation of how semantically-driven and rule-based diagrammatic reasoning were psychologically plausible for particular cases of scientific practice: an actual trackless path reconstruction in bubble chamber experiments, as reported by Galison (1997). I will propose "cognitive imagery projection and manipulation" (CIPM) as the most plausible psychological (perceptual/attentional/cognitive) mechanism matching the specific explanatory requirements for the case study, outlining the most significant current theories about this mental phenomenon (Shimojima 2011).

**Keywords:** Diagrammatic reasoning · Imagery · Particle physics

## 1 Introduction and Methodology

In this paper I aim to develop a cognitive evaluation of how semantically-driven and rule-based diagrammatic reasoning (departing from the experimental data of Shimojima [4]) could be psychologically plausible, for particular cases of scientific practice [1]: an actual trackless path reconstruction in bubble chamber experiments. The case study reported by Galison [2] will be briefly described in the next section. In Sect. 3, I will propose "cognitive imagery projection and manipulation" (CIPM) as the most plausible integrative psychological (perceptual/attentional/cognitive) mechanism, empirically tested by Shimojima [4] matching the specific explanatory requirements for this historical case study. I will flesh out in Sect. 4 every diagrammatic operation actually performed by "scanners" in a trackless path reconstruction of an invisible particle within a bubble chamber picture following the CIPM model.

## 2 Case Study: Visualizing Particles by Diagrammatic Operations

In the postwar period of 1954–1968, bubble chambers experiments were the forefront of scientific experimentation. Bubble Chambers consisted of a tank full of ionized superheated liquid, wherein every elastic or inelastic collision happening within the chamber, after a neutrino beam were emitted, were exhaustively photographed, generating uncountable kilometers of film containing bubble chamber pictures (BCP).

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 779–782, 2018.
https://doi.org/10.1007/978-3-319-91376-6_80

Woman "scanners" had to explore vast amounts of BCP in order to visually recognize plausible interesting events, just with their own cognitive abilities and a limited set of basic technical instruments.

Galison [2] reported that in some occasions "scanners learned to "see" the unseen by "reconstructing the paths of trackless neutral particles", for instance when two visible diverging curves appears ("V" shaped pattern) on a BCP with no visible source. Although invisible, scanner inferred that must actually exist an invisible particle causing the two visible "V" shaped curved tracks by means of a complex and extremely fast diagrammatic reconstruction of this invisible particle's trajectory. Although, Galison did not explained how was psychologically possible that scanners "learned to see the unseen" by means of trackless path reconstructions, as they in fact did. It seems reasonably and not trivial at all to understand *how* they did it, because, as I will methodologically defend, the production of evidential knowledge in this scientific practice strongly depends on developing highly-specific diagrammatic abilities. By understanding how those perceptual-cognitive processes underlying diagrammatic operations work, we will gain much comprehension on the role of diagrammatic reasoning in the general production of scientific knowledge.

## 3 Cognitive Imagery Projection-Manipulation Underlying Diagrammatic Operations

Here, I will postulate a psychological mechanism that will explain quite satisfactorily how scanners could identify/reconstruct invisible tracks of particles in BCP by means of (diagrams 1, 2 & 3). Cognitive Imagery Projection-Manipulation (**CIPM**):

**Cognitive Imagery:** CIS process information contained in Imagery States functionally-depended on their representational content (Imagistic Meaning) in psychological coordination with perceptual/attentional processing. **(Cognitive) Imagery Projection/Manipulation:** Cognitive imagery can be mentally projected either toward the visual scene or toward the physical scenery. Cognitive imagery can be mentally manipulated across the visual/physical scenery and processed by CIS in coordination with visual system.

The visual content already present within the sample BCP (e.g. the two diverging curved tracks) will remain fixed all along the reconstructive-inferential process. On the other hand, many imagistic representations will be mentally added or projected by scanners women onto their visual space/BCP space during the line-of-flight reconstruction: the center of each curved track, the radius between each center and the common vertex or the diagonal of the parallelogram. Once fixed on the visual/BCP scenery, all the projected imagistic representations would be "literally" seen by the scanners. A current theory in the market concerning CIPM mechanism is the "hypothetical drawing" of Shimojima [4]. He the idea that meaningful imagery could be literally projected and indexed onto the visual scenery in order to perform semantically-driven visual inferences. Following an in vitro-like methodology, Shimojima studied in two different experimental setting the experimental subject's eye movements when semantic rules were given to them in real practices of diagrammatic and spatial reasoning with no physical manipulation. For instance, two imagistic elements "A" and "B" were vertically displayed on a screen and it was uttered that the upper element B was lighter than the lower A; afterwards it was uttered that a third invisible element "C" were lighter than B. The vast majority of the subjects projected the imagistic element C up to B by entertaining the imagistic meaning of "upper is lighter". These tests showed very positive results, in the form of statistical correlation between eye tracking and semantically-driven imagistic indexing, for the very mechanism of CIPM itself.

## 4 Diagrammatic Cognition in Bubble Chambers Experiments

Now, I will tend to explain by means of CIPM how each diagrammatic operation were actually carried by those scanners:

### 4.1 Radius Location and Right Angle Rotation (Fig. 1 Left)

First of all, it was necessary to measure the curvature of both curved tracks. With enough training, material curvature template were not mandatorily required so scanner women would start to use "mentally projected" curvature template to accurately locate the center of each curved track in a much faster-efficient manner. Once the center was visually indexed in a non-necessarily filled space on the BCP, the radius of each track could be determined by mentally projecting a straight line between each indexed center and the visible vertex, visually fixing these lines on the BCP. Then, radius of each curved track (charged particle paths) were inversely proportional the length of the tangent lines obtained. These classic-mechanical and electromagnetic theoretical assumptions were deeply taken by scanners entertaining its "imagistic meaning", namely the diagrammatic representation of those theoretical concepts: if they would mentally rotate [3] each previously projected radius 90º toward their respective curved track, their vector momentum will be visually fixed and their magnitude inverse-proportionally established.

### 4.2 Translation of Tangents and Parallelogram Construction (Fig. 1 Center)

When scanners "could (literally) see" the momentum of each curved track, it was relatively easy to visually infer that the invisible track causing the two-particle decay will spatially coincide with the *vector addition of both momenta*. This is what we would call the "imagistic meaning" of conservation of momentum in a three particles elastic collision. For obtaining the vector addition by the well-known "parallelogram method" are required complex cognitive operations: to imaginary translate each projected momentum along the other momentum vector's length and viceversa, preserving vector direction and magnitude. If the two translated vectors accurately coincided on a vertex, the correctly constructed projected parallelogram would be indexically fixed on the BCP. Then, scanners could visualized momentum addition by mentally tracing a straight line between the two more distant vertexes.

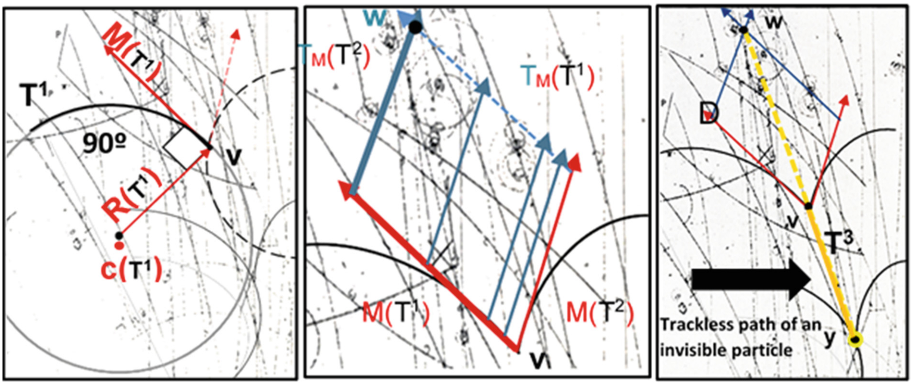


**Fig. 1.** Diagrammatic reconstruction of an invisible particle.

### 4.3 Diagonal Prolongation and Vertex Intersection (Fig. 1 Right)

In order to "visualize" the invisible line-of-flight, scanner women imaginarily (and indefinitely) prolonged the imagistically drown parallelogram's diagonal towards the lower vertex direction. If this projected line-of-flight intersect at some point an "interesting" pattern of tracks, for instance other two diverging curved tracks, then this might plausibly be the source vertex-represented event from where the neutral particle where emitted until it collided in the initial vertex/event, causing the two-particles decay. In this case, the invisible segment or line-of-flight will be de-limited by two vertex already present in the BCI.

## 5 Conclusion

Beside almost all recent studies on diagrammatic cognition are focused on every day cases or simple experimental cases, I have tried to show in this study that it is also possible to understand the complex psychology underlying cases of diagrammatic reasoning in real scientific practice as the one introduced along this paper. For carrying such operations, were necessary to project and manipulate cognitive imagery in an extremely efficient and accurately manner. In conclusion, diagrammatic cognition had constantly played a deeply disregarded but highly decisive role both in the development of leading experimental high-energy physics.

## References

1. Carruthers, P., Stich, S., Siegal, M. (eds.): The Cognitive Basis of Science. Cambridge University Press, Cambridge (2002)
2. Galison, P.: Image and Logic: A Material Culture of Microphysics. University of Chicago Press, Chicago (1997)
3. Hegarty, M.: Mechanical reasoning by mental simulation. Trends Cogn. Sci. **8**(6), 280–285 (2004)
4. Shimojima, A.: Operational constraints in diagrammatic reasoning. In: Allwein, G., Barwise, J. (eds.) Logical Reasoning with Diagrams. Oxford University Press, Oxford (1996)

# Teaching Argument Diagrams to a Student Who Is Blind

Marc Champagne(✉)

Trent University, Peterborough, ON K9L 0G2, Canada
thinkingmarcchampagne@gmail.com

**Abstract.** This paper describes how bodily positions and gestures were used to teach argument diagramming to a student who cannot see. After listening to short argumentative passages with a screen reader, the student had to state the conclusion while touching his belly button. When stating a premise, he had to touch one of his shoulders. Premises lending independent support to a conclusion were thus diagrammed by a V-shaped gesture, each shoulder proposition going straight to the conclusion. Premises lending dependent support were diagrammed by a T-shaped gesture, the shoulder premises meeting at the collar bone before moving down to the belly button. Arguments involving two pairs of entailments were diagrammed by an I-shaped gesture, going from the collar bone to a mid-way conclusion above the abdomen before travelling to the final conclusion at the belly button. The student's strong performance suggests that placing propositions at different locations on the body and uniting them with gestures can help one discern correct argumentative structures.

**Keywords:** Arguments · Body · Gestures

Last year, I did something I had never done before: I taught argument diagramming to a student who cannot see. Shortly before the start of classes, I was informed by my university's Accessibility Services that one of my students was blind and would therefore need accommodations. This posed a problem, since I usually spend the first month of my Critical Thinking course teaching students how to convert argumentative texts into diagrams with numbers and arrows. It was made clear to me that, given the costs and delays involved, Braille was not an option (even with unlimited resources, Braille struck me as an inferior option). A puzzle-like device has been built to help blind students figure out categorical syllogisms [1], but syllogisms are only one type of argument, while I was aiming for greater generality. I could have assigned different course content. However, I had just published an article arguing that non-visual logics are possible [2]. My student's needs let me put this academic background to good use.

Diagrams are signs that mimic only relations, not relata [3]. Consider the argument diagrams taught in Critical Thinking courses. Typically, individual propositions are represented by numbers. Although this numerical assignment is arbitrary, the skeletal structure that emerges when one relates the numbers is not, since that structure is answerable to the logical relations holding among the various propositions. I was already committed in print to the idea that diagrams can express such relations non-visually, so I liked the idea of now putting that commitment to the test.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 783–786, 2018.
https://doi.org/10.1007/978-3-319-91376-6_81

Research has shown that, “[i]f the environment in which learning occurs is not supportive to students with visual impairments, their learning will automatically be interrupted” ([4], pp 22–23). I therefore met with my student beforehand, at the start of the term. When he came to my office, I could sense that he was nervous. I therefore broke the ice: “Listen, I’m going to be frank. I have never taught argument diagrams with any channel other than vision. I have some promising ideas about how we might do this. But, I want to be clear from the get-go that this will be a first, for both of us.” This transparency put him at ease. The epistemological fallibilist in me then stepped in: “Let’s start from the assumption that we are going to mess up. If we indeed mess up, then I will give you a month-long extension, so that you can complete a different chapter from our textbook. If, however, things work and you end up acquiring the skill in question, there will be no need for an extension. Agreed?” He agreed.

Our first session was devoted to introducing the method. The goal was to translate a string of symbols, namely a textual passage, into a diagram that exhibits the argumentative structure at play in that passage. My student showed me how he uses a computer program to hear (in a monotone robotic voice) what is written. Importantly, he has to remember all that he hears. Since there are limits to what one can recall, I decided to make adjustments. First, I made sure that the arguments would be at most three propositions long (we were going to spend a mere four weeks on this, so we had to pick battles big enough to matter but small enough to win). Second, I typed all the passages in a simple Text document, since I discovered that extra formatting merely hindered his computer reader. Third, I made sure to skip a line after each sentence. The computer program did not make any noticeable pause when it encountered a period, so I wanted to give my student a chance to really bite down on these grammatical units. Fourth, I typed out the exercise numbers with regular language, since the computer program reads numbers 0 to 9 quite well but starts saying things like “One Six” when it encounters larger numbers like “16.” Finally, I lifted all time constraints and allowed my student to listen to auditory contents as often as he wished.

The opening drill consisted in listening to a variety of arguments and picking out the main conclusion. What is the point being made? My student had to state this conclusion out loud. I gave him immediate feedback after each answer. To supply him with exercises, I usually brought a USB key with a Text file that he copied onto his laptop. However, for this warm-up exercise, I read him the arguments from a book, repeating them in a clear voice whenever he wished.

The second task was similar to the first, only this time my student had to pick out the reasons given to support the conclusion. I noticed that he had a tendency to shorten the sentences by rewording them. I warned against doing this, since important parts connecting the premises risk being discarded.

In the next task, we started incorporating the body. The goal was to exploit my student’s prior familiarity with his own lived body in order to structure the layout of arguments. When stating a conclusion out loud, he had to touch his belly button. When stating a premise, he had to touch one of his shoulders. Since I could see him making these gestures, he no longer had to preface his answers with verbal locutions like “The conclusion is…” and “The first premise is…” Because we were working with a maximum of three propositions, there was always enough space on his torso to locate the relevant constituents of an argument.

At this early stage, I included only two types of arguments, namely those where the premises lend dependent support and those where they lend independent support. Once my student has pinned the conclusion to his belly button, he would determine the type of support by going from one shoulder to the belly and asking whether that relation made any sense. Independent relations of support were thus signified by a V-shaped gesture where each shoulder proposition went straight to the conclusion. An example of an independent argument would be “Sushi is made from seaweed. Sushi is made from rice. Therefore, sushi contains plant matter.” Dependent relations of support were signified by a T-shaped gesture that first connected the premises at the collar bone before moving straight down to the belly button. An example of a dependent argument would be “Either the baby in my belly is kicking or I have gas. I do not have gas. Therefore, the baby is kicking.” As my student announced the propositions and performed the gestures, he looked like a Christian making the sign of the cross.

For every argument, I required my student to move in two directions before settling on his answer. When working downward from the shoulders to the belly button, he had to use connecting words like “and” and “Therefore,” timing those verbal cues with his gestures. When moving upward from his belly button to his shoulders, he had to use transition words like “Why? Because…” If there was an error in his preliminary diagram, it tended to get exposed when it was tested in both directions.

After three weeks of practice, I felt that my student was ready for his first quiz. This quiz was comprised of 10 questions, each worth 1%. He had his headphones on, so I was unable to hear how many times he re-listened to the texts on his computer. I think this privacy was a good thing, since it left him free to listen to the arguments as many times as he needed, without worrying about what I might think. While he deliberated, I worked silently on something else (I told him as much). My student tended to do his trial gestures silently. However, I think that talking out loud would have been better, pedagogically speaking. In any event, when he was done toying with an argument, he notified me that he was ready to give his answer. I would then look and listen as he diagrammed the argument. I graded his results silently on a notepad, scoring his answers in an all or nothing manner, giving 1 point for a perfect diagram and 0 for anything else. I was able to divulge his quiz result immediately after he was done.

My student was clearly intent on doing well, so he was putting a lot of pressure on himself. In previous sessions, I noticed that he would occasionally rush to a verdict. Diagrammatic reasoning is at its most fertile when it includes an element of play, so I encouraged my student to toy creatively with the different logical-*cum*-gestural relations, feeling out which fits best.

The weeks of in-office and at-home practice proved sufficient, since my student earned 10 out of 10 for his quiz. We kept practicing for a more challenging test that would be worth 25%. This test added a new diagram to the mix, namely chain arguments. In this structure, a premise (starting at the collar bone) leads to a mid-way conclusion (located above the abdomen) which in turn serves as a premise for a final conclusion (at the belly button). If dependent support is T-shaped and independent support is V-shaped, then this iterated relation of entailment yields a bodily diagram that is I-shaped. An example of a chain argument would be “Sarah must go visit her parents for the holidays. As a result, she will need to take the train. Therefore, she will need money.” What I learned was that, on its own, this diagrammatic structure was easy to grasp, no doubt because it reduces the threefold linkage to two pairs. That said, when chain arguments were randomly mixed with dependent and independent arguments, they became harder to spot. My student scored 20 out of 25 for this test.

All told, we worked on bodily diagramming for a month before moving on to the next course content. One promising possibility we had no time of exploring is that of recycling a conclusion into a premise, by sliding it from the belly button to a shoulder. I suspect that, once mastered, this could allow one to handle lengthier arguments.

In any event, the foregoing has been predicated on a realistic, not an idealized, conception of how inquiry proceeds. Innovations emerge from the vicissitudes of practical engagements (that are often best expressed in a narrative format). While the positive learning/teaching experience I have recounted was insufficiently controlled to justify full-on theses about diagrammatic reasoning and argumentative cognition, my student's strong performance suggests that placing propositions at different locations on the body and uniting them with gestures can help one discern correct argumentative structures.

Although the diagramming method that I have described was designed to meet special learning needs, I see no reason why it could not be used in regular class contexts. The next step, then, would be to teach this method to more students (as a main skill or side drill), make adjustments where necessary, and report back the findings. Teachers and students are well placed to judge whether a given exercise works, so ideally their first-hand reports should be incorporated into experimental designs [5].

Given that there are always students who recoil from mainstream notations, it might be worthwhile to have an alternative method on stand-by that appeals to a sense modality other than sight. There are many ways to learn [6] and many ways to make an argument [7], so hopefully this bodily diagramming can add a useful arrow to the teacher's quiver.

## References

1. Goldstein, L.: Adding a dimension to logic diagramming. In: Blackburn, P., van Ditmarsch, H., Manzano, M., Soler-Toscano, F. (eds.) TICTTL 2011. LNCS (LNAI), vol. 6680, pp. 101–108. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21350-2_13
2. Champagne, M.: Sound reasoning (literally): Prospects and challenges of current acoustic logics. Logica Universalis **9**, 331–343 (2015). https://doi.org/10.1007/s11787-015-0127-7
3. Peirce, C.S.: Collected Papers. Harvard University Press, Cambridge (1931–1958)
4. Mwakyeja, B.M.: Teaching students with visual impairments in inclusive classrooms. Master's thesis, Faculty of Educational Sciences. University of Oslo, Oslo (2013). https://www.duo.uio.no/bitstream/handle/10852/36642/MasterxsxThesis.pdf?sequence=1
5. Gallagher, S., Sørensen, J.B.: Experimenting with phenomenology. Conscious. Cogn. **15**, 119–134 (2006). https://doi.org/10.1016/j.concog.2005.03.002
6. Gardner, H.: Frames of Mind: The Theory of Multiple Intelligences. Basic, New York (2011)
7. Groarke, L.: Going multimodal: What is a mode of arguing and why does it matter? Argumentation **29**, 133–155 (2015). https://doi.org/10.1007/s10503-014-9336-0

# Correction to: How Cross-Representational Signaling Affects Learning from Text and Picture: An Eye-Tracking Study

Juliette C. Désiron, Mireille Bétrancourt,
and Erica de Vries

## Correction to:
## Chapter "How Cross-Representational Signaling Affects Learning from Text and Picture: An Eye-Tracking Study" in: P. Chapman et al. (Eds.): *Diagrammatic Representation and Inference*, LNAI 10871, https://doi.org/10.1007/978-3-319-91376-6_68

By mistake, the original version of this chapter was not published open access. The publishing mode has been changed to open access.

---

The updated version of this chapter can be found at
https://doi.org/10.1007/978-3-319-91376-6_68

© Springer International Publishing AG, part of Springer Nature 2019
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, p. C1, 2019.
https://doi.org/10.1007/978-3-319-91376-6_82

# The Role of Diagrams in Contemporary Mathematics: Tools for Discovery?

Jessica Carter

University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark
`jessica@imada.sdu.dk`

**Abstract.** My presentation draws upon and combines results from some of my articles on the role of diagrams in contemporary mathematics. Referring to [1, 2] I will present examples of how diagrams function as tools for discovery in contemporary analysis. The purpose in this talk is to analyse why these diagrams are fruitful. In most of the talk I use 'diagram' in its ordinary sense referring to certain (2-dimensional) visual representations composed of lines and sometimes letters standing for mathematical objects.

**Keywords:** Role of diagrams · Contemporary mathematics · Fruitfulness

Carter [1] considers a case study from free probability theory (a field combining analysis with probability theory) where certain diagrams are used to represent permutations (and constructions on permutations). The diagrams give rise to the concepts of a 'crossing permutation' (Fig. 2) and a 'neighbouring pair' (Fig. 1). That is, new concepts are found by representing permutations by diagrams. Furthermore these diagrams can be "manipulated" and by doing this new results have been discovered. One of these results establishes a connection between the two concepts, that is, of a crossing permutation and a permutation having no neighbouring pairs. Note that these properties can be *shown* in a diagram – in a crossing permutation lines *cross* and neighbouring pairs are *neighbouring numbers.* That is, in addition to giving rise to new concepts, the diagrams visualise these concepts as well as relations holding between them.

Carter [2] discusses an example where manipulations with diagrammatic presentations of directed graphs have contributed to establish that a certain type of C*-algebras exists. In the theory of C*-algebras an important question concerns their classification, that is, determining which different types of algebras exist up to isomorphism. An important tool in order to do this is to compute their K-groups, denoted $K_0$ and $K_1$. Unfortunately these K-groups are difficult to define. Recently an easier way has been found by instead generating C*-algebras and their corresponding K-groups from so-called directed graphs. A directed graph consists of a collection of vertices and directed edges between these vertices. See Fig. 3 for an example.

A particular graph gives rise to a collection of generators and relations from which a C*-algebra may be defined. Read in a different way the graph gives rise to a linear map. From this map one can easily define the two groups, $K_0$ and $K_1$. One natural question that arises in this context is how many different C*-algebras can be obtained

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 787–790, 2018.
https://doi.org/10.1007/978-3-319-91376-6

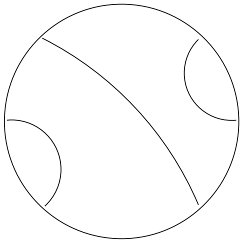


**Fig. 1.** A representation of a permutation with neighbouring pairs, (1,2) and (4,5).

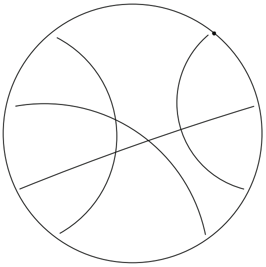


**Fig. 2.** A representation of a crossing permutation.

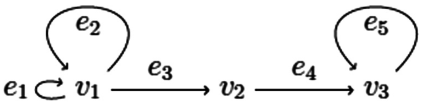


**Fig. 3.** A picture of a directed graph with vertices, $v_1, v_2, v_3$ and edges $e_1, e_2 \ldots e_5$. Arrows show the direction of the edges.

from directed graphs. It is a response to this question that is considered in [2]. The result is found by *manipulating* the directed graphs while calculating their associated K-groups.

In both of the above cases diagrams function as objects – or signs – that can be manipulated and so experimented on. Furthermore I noted in the first case study that certain properties and relations are directly observable in diagrams. These two features are central to Peirce's description of 'diagrammatic reasoning' and so in the second part of my talk I will give a few relevant details of what he means by this (referring to [3–5]). First it should be noted that according to Peirce a 'diagram' includes representations that one would normally not consider as diagrams. Moreover it should be stressed that it is not entirely clear what Peirce took a diagram to be. Peirce states that a diagram is an icon and in some places a representation of *relations* or *rationally related objects* see ([6], pp. 316–317). According to Peirce diagrammatic reasoning is a description of (mathematical) necessary reasoning. Thus an important role of a diagram is to allow you to see the necessary relation holding between the antecedent and conclusion of a proposition. Reasoning proceeds by constructing "*a diagram, or visual* *array of characters or lines. Such a construction is formed according to a precept furnished by the hypothesis. Being formed, the construction is submitted to the scrutiny of observation, and new relations are discovered among its parts, not stated in the precept by which it was formed, and are found, by a little mental experimentation, to be such that they will always be present in such a construction*"([7], CP 3.560). This seems to fit the procedure in Euclid's Elements. A diagram is constructed so that it represents the hypothesis of the stated proposition. Sometimes further constructions have to be made (e.g. lines drawn) until the conclusion can be seen to follow from the diagram. What is remarkable is that Peirce holds that this characterises mathematical reasoning in general (and so accordingly he extends the notion of 'a diagram'). To mention a simple example, a proof consisting of a calculation to prove the proposition *the product of the sum and difference of two numbers is equal to the difference between their squares* would also be a diagrammatic proof. It is important for Peirce that a proof – or the drawn diagram – is a concrete representation, in other words a token, so that it is possible to *observe*. At the same time the diagram is the representation of a symbolic statement (or the interpretant of a symbolic statement) and so general. The combination of the two gives the necessity of the conclusion.

Peirce's idea of diagrams as concrete objects that can be *experimented on* and *observed* fits well with the observations made in the two case studies above. Firstly, in both cases are certain representations that were manipulated with when discovering new results. The idea of a sign that can be manipulated gives rise to the notion of a 'faithful representation', which I propose may be used to explain the fruitfulness of certain representations [2, 8]. Note that a faithful representation can be any type of representation, including a formal expression, in case it fulfils the characterisation given below.

A faithful representation fulfils that (i) it resembles or shares certain relations with the represented object (i.e., represents iconically) and (ii) manipulations can be performed on the representations, respecting relevant relations between the represented objects, so that new relations may become visible.

Secondly I noted in the first case study that certain concepts as well as relations between them are shown in the diagrams. (Now taking 'diagram' in its usual meaning.) In both case studies it is possible to compare diagrammatic representations with formal representations of the same concept. Here the diagrams *show* the relations holding whereas the formal expressions *describe* them. Furthermore in both cases it can be argued that the diagrammatic presentations offer a cognitive advantage over formal expressions. In the case study on graph algebras, for example, a case can be made that it is easier to read off relevant information from the diagrammatic presentation of a graph. I measure 'cognitive advantage' in terms of the number of cognitive resources drawn upon in order to perform a given task. In conclusion diagrams seem to offer two advantages that I propose are relevant components in a characterisation of understanding. First is their capacity in some cases to *show* relations in contrast to describing them. Second is the fact that they sometimes offer cognitive advantages.

## References

1. Carter, J.: Diagrams and proofs in analysis. Int. Stud. Philos. Sci. **24**(1), 1–14 (2010)
2. Carter, J.: Graph-algebras – faithful representations and mediating objects in mathematics. Under review in Endeavour
3. Carter, J.: Logic of relations and diagrammatic reasoning: structuralist elements in the work of Charles Sanders Peirce (1839–1914). In: Reck, E., Schiemer, G. (eds.) The Prehistory of Mathematical Structuralism. OUP (to appear)
4. Marietti, S.: Observing signs. In: Moore, M.E. (ed.) New Essays on Peirce's Mathematical Philosophy, pp. 147–167. Open Court, Chicago and LaSalle (2010)
5. Stjernfelt, F.: Diagrammatology. An Investigation on the Borderlines of Phenomenology, Ontology, and Semiotics. Springer Science and Business Media (2007)
6. Peirce, C.S., Carolyn, E. (eds.): The New Elements of Mathematics, vol. IV. Mouton, The Hague (1976)
7. Peirce, C.S.: Collected Papers of Charles Sanders Peirce, vol. I–IV. (First printing 1931–1933, third printing 1965–1967.) Edited by C. Hartshorne and P. Weiss. Belknap Press of Harvard University Press, Cambridge (1965–1967)
8. Carter, J.: Exploring the fruitfulness of diagrams in mathematics. Synthese. https://doi.org/10.1007/s11229-017-1635-1

# Unveiling Darwin's Theory of Evolution Through the Epistemological Study of His Diagram

Fabien Ferri

Logiques de L'agir, University of Burgundy Franche-Comté,
EA 2274 Dijon, France
fabien.ferri@univ-fcomte.fr

**Abstract.** The expression of scientific knowledge relies on a set of non-linguistic representations: "pictorial representations", such as encyclopaedic illustrations (photographs, paintings, figurative drawings, etc.), "schematic representations", such as textbook illustrations (for instance the drawing of a cell in a biology textbook) and finally "diagrammatic representations" [1, Chap. 7], such as those found in subjects like physics which are used to simplify calculatory expressions (Feynman diagrams, Penrose diagrams, etc.). Diagrammatic representations are of particular interest because they play a crucial role, which is not only pedagogical and heuristic, but also epistemic. This paper will endeavour to contribute to the study of the epistemic role of diagrammatic representations through a notable and yet little-noticed case-study: Darwin's diagram. Introduced in the middle of the fourth chapter of his *On the Origin of Species* (1859), this diagram could at first be mistaken for a family tree. Through an analysis of all the elements of the diagram (continuous lines, dotted lines, letters, Roman numerals, etc.) and Darwin's own comments, this paper will show how the explanatory hypotheses of a major scientific theory can be best understood thanks to this tool of epistemic representation.

**Keywords:** Darwin's diagram · Epistemic representation · Theory of evolution

## 1 Introduction

Darwin's diagram is well-known as it has been cited many times, both by science historians and biologists since the 1859 publication of *On the Origin of Species* [2]. In spite of this, it seems that a number of features have been neglected. This figure is unique because it is the only non-linguistic representation used in the book: it is introduced and commented on in Chap. 4, then reinterpreted by Darwin in Chaps. 10 and 13, either to emphasize its vertical dimension or to emphasize its horizontal dimension. Why did Darwin choose to call this figure "diagram"? What does it represent? A classification? A family tree? I will show that the "queer diagram" [3, p. 123] is indeed a diagrammatic representation serving an epistemic function whose object is Darwin's theory of evolution. Moreover, I will demonstrate that it is a "theoretical representation" [1, p. 367] insofar as it expresses hypotheses and exhibits mechanisms, which grants it predictive and explanatory power.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 791–795, 2018.
https://doi.org/10.1007/978-3-319-91376-6

## 2 Symbolic System and Diagrammatic Representation

Before answering these questions, it is necessary to briefly return to the analyses initiated by Nelson Goodman [4]. Goodman, through the notion of "symbolic system", sought on the one hand to generalize the notion of language in order to extend it to non-linguistic representations, and on the other hand, to classify these representations into different types: "pictorial representations" (which look like their objects), "schematic representations" (which keep the topological structure of their target) and "diagrammatic representations" (which spatially represent non-spatial relationships) [1, Chap. 7]. A symbolic system can thus be defined as a set of marks that first maintain syntactic relations between each other, but also semantic relations with the objects of a field of reference, in order to fulfil various functions; be they aesthetic, mnemotechnic or epistemic. This paper will study the epistemic function of diagrammatic representations.

The diagram (see Fig. 1) illustrates the mechanism of natural selection by highlighting four principles (variation, descent, divergence and extinction) that converge towards the same phenomenon: descent with modification. How is it organized?

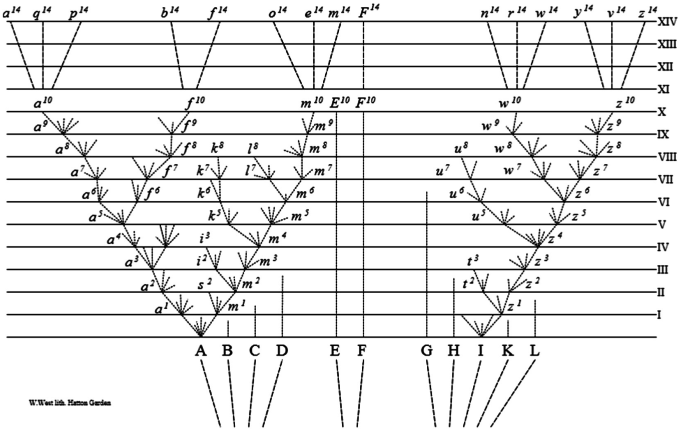


**Fig. 1** Darwin's diagram. Charles Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, 1st ed., London, John Murray, 1859, Chap. 4, p. 116–117.

## 3 Epistemological Analysis of Darwin's Diagram

First of all, the diagram is organized around two axes: one vertical, the other horizontal. It reads from bottom to top. On the vertical axis, a series of 14 Roman numerals indicate the course of time: this axis does not measure duration, but stratifies generations. It can be interpreted as a generational axis (this is actually what Darwin does in Chap. 4) or as a paleontological depth axis (as Darwin does in Chap. 10). On the horizontal axis, a series of 11 upper case letters represent 11 varieties, which can refer to different species as well as varieties of the same type, or variations within a single population or even several populations, according to Darwin's comments. This enables him to "quantify freely", as Grothendieck argued when discussing the theory of categories [5, p. 259]. It was the mathematical theory of category that directly inspired contemporary biology since it tried come up with more natural formalisms able to model living beings [6].

However, it is essential to understand that Darwin's aim is not to expose a phylogenetic classification of living forms, but to symbolize a "dangerous idea" [7]. According to this idea, there cannot be any intelligent design to explain the principle of the evolution of species, but rather a process of natural selection without purpose, which operates when chance and necessity work together, in a Malthusian context of struggle for existence [8, p. 21].

Moreover, the 11 letters below the first horizontal line express a morphological, anatomical and ecological divergence: their variable deviation stands for the relative distance between varieties or species (it can be noted that there is more proximity between *E* and *F* than between *F* and *G*). This means that the uppercase letters represent undetermined taxonomic categories: they are indeterminate variables that allow taxonomic categories to be freely specified. By proceeding this way, Darwin does exactly what the philosopher of sciences Jean Cavaillès calls an operation of "thematization", a notion whose paternity seems to go back to Husserl [9, p. 165]: Darwin indeed turns properties into objects whose properties he will study. Variation, filiation, divergence and extinction are properties that are objectified by the lines of the diagram and polarized following the two axes: this is what allows "taxonomic branches" to appear [3, p. 127]. In allowing the survival and extinction of species to be perceived in the blink of an eye, Darwin captures the infinite variability of nature in a finite synoptic representation: the diagram thus symbolizes a dynamic totality through a static representation [10, Chap. 5]. However, it should be noted that it was because of a competitive context that Darwin was forced into representing his theory of evolution in a diagram that was inspired by the image of the tree. Indeed, had Alfred R. Wallace not rehabilitated this crucial image, Darwin might have not have been inclined to do so himself [10, Chap. 4].

If the lines of each close variety (*A*, *B*, *C*, *D*) are extended, they can be connected outside the diagram to converge towards a point of origin that allows the hypothesis of a group or population of origin to be formulated. Conversely, the dotted lines that run from *A* to the upper strata do not represent the actual descent of this variety, but the resulting variations. It can be noted that there is a high variability for population *A* whereas there is none for population *D*. The more the initial variations, the more the population *A* proliferates over time: the differences, which were first minimal, lead to considerable differentiation in the long run. As time goes by, the populations differ from one another. The diagram thus performs a synoptic function, because it allows to present simultaneously different time perspectives.

Darwin then displays new lowercase letters to articulate this differentiation (*a*, *m*, *z*, etc.). If the horizontal lines numbered from I to X indicate the variations of each generation, they only continue one of them when it gives rise to a major deviation. On the last four horizontal lines (XI to XIV), Darwin draws general evolutionary lines ($a^{14}$, $q^{14}$, $p^{14}$, etc.) that disregard the variations abandoned during evolution (between lines I and X).

Thus, Darwin succeeds in showing that the same starting population can be divided into different species, families, types and classes very distant from each other and that are no longer tied to the former population. After fourteen temporal sequences, *A* no longer exists: it turned into eight new varieties more or less distant from each other ($a^{14}$, $q^{14}$, $p^{14}$, $b^{14}$, $f^{14}$, $o^{14}$, $e^{14}$, $m^{14}$). Conversely, other species, such as *E* and *F*, have no variation: unchecked and unmodified, they can withstand time without diversifying. On the other hand, some lines of populations, for example those of *C* or *D*, meet over time with the lines of other populations that have proliferated: if they were prolonged, they would come into conflict within the drawing. Darwin thus thereby illustrates the competition between species. Better adapted, the new varieties supplant the old ones, which then go extinct. This means that varieties that are too close compete with each other. Therefore, the more variability in a population and the more divergent the characteristics, the more likely a population is to survive, proliferate and diversify.

The forms of stratum XIV may no longer be able to interbreed or have descendants together, but they share a common ancestor if their line is extended, which explains why life forms can keep common characteristics that they inherited from extinct ancestors.

## 4 Conclusion

Darwin's diagram is a diagrammatic representation in the strong sense of the word, because the spatial relationships it represents do not represent spatial relations as in pictorial or schematic representations, but causal, temporal and hierarchical relations between life forms. Moreover, if its object is the theory of evolution, the object is actually an abstract set of mechanisms: for example, the intergenerational genealogical mechanism or the mechanism of coordination of variation and heredity leading to divergence and extinction. Its elements, and the way in which they are organized, allow us to express theoretical hypotheses: such as this imaginary prediction made by Darwin in the first edition, but withdrawn from the second, according to which the power of natural selection could, with time, transform a bear into a whale [3, p. 135]. (Translated by Lucie Lopez)

## References

1. Vorms, M.: Théories, modes d'emploi: une perspective cognitive sur l'activité théorique dans les sciences empiriques. Université Panthéon-Sorbonne – Paris 1 (2009)
2. Darwin, C.: On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, 1st edn. John Murray, London (1859)
3. Hoquet, T.: Darwin contre Darwin: Comment lire L'Origine des espèces. Seuil, Paris (2009)
4. Goodman, N.: Languages of Art: An Approach to a Theory of Symbols, 2nd edn. Hackett Publishing, Indianapolis (1976)
5. Krömer, R.: La théorie des categories: ses apports mathématiques et ses implications épistémologiques. Université Nancy 2 and Universität des Saarlandes (2004)
6. Varenne, F.: Théorie mathématique des categories en biologie et notion d'équivalence naturelle chez Robert Rosen. Revue d'histoire des sciences **66**(1), 167–197 (2013)

7. Dennett, D.C.: Darwin's Dangerous Idea: Evolution and the Meanings of Life. Simon & Schuster, New York (1995)
8. Ragan, M.A.: Trees and networks before and after darwin. Biol. Direct **43**(4), 1–38 (2009)
9. Mélès, B.: Pratique mathématique et lectures de Hegel, de Jean Cavaillès à William Lawvere. Philosophia Scientiae **16**(1), 153–182 (2012)
10. Bredekamp, H.: Darwins Korallen. Die frühen Evolutionsmodelle und die Tradition der Naturgeschichte, Verlag Klaus Wagenbach, Berlin (2005)

# Can a Picture Prove a Theorem? Using Empirical Methods to Investigate Diagrammatic Proofs in Mathematics

Josephine Relaford-Doyle and Rafael Núñez

University of California – San Diego, La Jolla, San Diego, CA 92093, USA
`jorelafo@ucsd.edu`

**Keywords:** Diagrammatic proof · Mathematical reasoning
Philosophy of mathematics

## 1 Background

The necessity of results obtained through formal proof is a defining feature of modern mathematics, and so it is not surprising that the question of what constitutes a valid proof is of central concern in the philosophy of mathematics.Currently, the dominant view is that a proof must consist of a series of propositions in which each statement follows logically from those preceding. Diagrams, while widely recognized as useful heuristic devices, are generally not considered an acceptable means of mathematical proof. Recently, however, various scholars have argued for an expanded role for visual representations in mathematical justification [1], with some going so far as to claim that "pictures can prove theorems" [2].

The claim that a diagram can function as a proof is controversial for a number of reasons [3].Most notable for the study reported here is the concern that diagrammatic proofs are necessarily finite and particular, while the theorems they are intended to prove generally encompass an infinite number of cases. Nevertheless, diagrammatic proofs have been described as "rapidly and deeply convincing", in some cases even more so than their formal counterparts [4].

To illustrate these issues, consider the theorem $1 + 3 + \ldots + (2n - 1) = n^2$. In Fig. 1(a) we verify that this theorem is true for the $n = 1$ through $n = 4$ cases, which might lead us to conclude (via inductive reasoning) that the general theorem is *likely* true. However, these examples alone do not constitute a proof that the theorem is *necessarily* true for *all* natural numbers; this would require a formal proof (in this case, a proof by mathematical induction would work). Fig. 1(b) shows a diagrammatic proof of the same theorem. The diagram only displays the first four cases of the theorem; however, the image contains structure that may suggest that the pattern will continue to hold as new layers are added. Indeed, diagrammatic proofs like the one in Fig. 1(b) have been described as "completely convincing" [5]. Moreover, unlike a formal proof by mathematical induction, the diagrammatic proof is free of complex notation and sophisticated algebra; successful interpretation seems to rely only on simple concepts like odd numbers, addition, and squaring. The apparent simplicity of the image has led

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 796–799, 2018.
https://doi.org/10.1007/978-3-319-91376-6

some to suggest that this diagram can function as a proof even for viewers with only "basic secondary school math knowledge" [6], and for viewers who are not familiar with formal mathematical induction [2].

(a) $1 = 1^2$

$1 + 3 = 4 = 2^2$

$1 + 3 + 5 = 9 = 3^2$

$1 + 3 + 5 + 7 = 16 = 4^2$

(b)

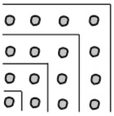


**Fig. 1.** Numerical (a) and diagrammatic (b) evidence that $1 + 3 + \ldots + (2n - 1) = n^2$.

In this study, originally reported in [7], we applied empirical methods to investigate the extent to which this diagram may serve a proof-like function for viewers. Specifically, we asked two questions: (1) Given the diagrammatic proof, do viewers generalize the theorem to cases not depicted in the image? (2) If so, is that conclusion considered necessary, as in formal mathematical proof, or only likely, as in standard inductive reasoning? To explore claims that the image is interpretable even by viewers with no particular mathematical training, we also assessed the effect of expertise with formal proof-writing on interpreting the image.

## 2 Methods

We recruited participants from two distinct populations. Our first group of participants (n = 25) was drawn from the general subject pool and consisted of university-level students with a variety of majors including psychology, cognitive science, and linguistics. None of these students had taken a university-level mathematics course on proof-writing, and so we call these participants "proof-untrained". Despite not having taken a course on proofs, these participants were all highly-educated adults enrolled at a prestigious university such that we expected them to have a firm grasp of the relatively simple mathematical concepts involved in the diagram. We recruited our second group (n = 24) of participants through the mathematics department; specifically, we enrolled participants who had received at least a B- in "Mathematical Reasoning", an upper-division mathematics course that covers a variety of formal proof strategies including mathematical induction. We refer to this group as "proof-trained".

We gave each participant the diagrammatic proof in Fig. 1(b) and asked him or her to explain how the image was related to the mathematical theorem it was intended to prove. We then conducted a semi-structured interview with each participant, which was designed to determine whether they had generalized the theorem and, if so, whether this conclusion carried the necessity associated with formal proof. We assessed generalization by asking each participant two questions: (1) "Do you think the statement is true in all cases?" and (2) "What would be the sum of the first 8 odd numbers?" Any participant who answered "yes" and "64" was classified as having generalized the theorem. To assess necessity, we then raised the possibility of large-magnitude counterexamples: “very large numbers where the statement actually isn’t true.” We gauged each participant’s resistance to this possibility on a 0–5 scale, where 0 indicated no doubt that counterexamples exist and 5 indicated complete rejection (stating counterexamples are impossible, indicating high mathematical necessity).

## 3 Findings

Participants in both groups generalized the theorem to cases not depicted in the image (Fig. 2, left); however, proof-untrained participants subsequently showed significantly less resistance to the possibility of counterexamples than did the proof-trained participants (Fisher Exact test, p = 0.008; Fig. 2, right). We take these results to suggest that for most participants without knowledge of formal proof-writing the diagram functions as set of a examples which allows for a standard inductive generalization, but does not provide the necessity associated with formal proof.

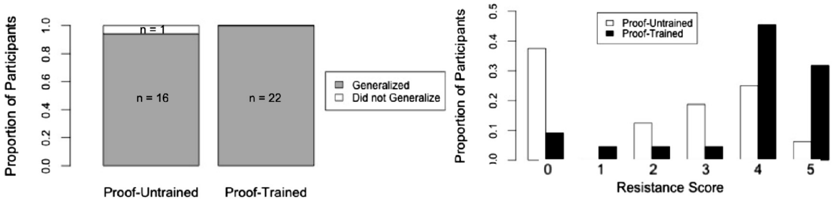


**Fig. 2.** Participants in both groups generalized the theorem (left), but only proof-trained participants showed high resistance to the possibility of counterexamples (right). Adopted from [7].

Thus, while the diagrammatic proof was in some sense “convincing” to all viewers, careful empirical work revealed nuances in the conclusions that viewers drew from the image. Specifically, the diagrammatic proof did not reliably convince proof-untrained viewers that the theorem was *necessarily* true for *all* natural numbers – this conclusion appears to rely on familiarity with formal proof strategies.

## References

1. Giardino, V.: Intuition and visualization in mathematical problem solving. Topoi **29**(1), 29–39 (2010)
2. Brown, J.R.: Philosophy of Mathematics: A Contemporary Introduction to the World of Proofs and Pictures. Routledge, New York (2008)
3. Hanna, G., Sidoli, N.: Visualisation and proof: a brief survey of philosophical perspectives. ZDM **39**(1–2), 73–78 (2007)
4. Doyle, T., Kutler, L., Miller, R., Schueller, A.: Proofs without words and beyond. Convergence. http://www.maa.org/press/periodicals/convergence/proofs-without-words-and-beyond. Accessed 3 Dec 2017

5. Chihara, C.S.: A Structural Account of Mathematics. Oxford University Press, New York (2004)
6. Jamnik, M.: Mathematical Reasoning with Diagrams. University of Chicago Press (2001)
7. Relaford-Doyle, J., Núñez, R.: When does a 'visual proof by induction' serve a proof-like function in mathematics? In: Gunzelmann, G., Howes, A., Tenbrink, T., Davelaar, E.J. (eds.) In: Proceedings of the 39th Annual Conference of the Cognitive Science Society, pp. 1004–1009. Cognitive Science Society, London (2017)

# Otto Neurath's Isotype and C. K. Ogden's Basic English

James McElvenny[1,2]

[1] University of Potsdam, Potsdam, Germany
james.mcelvenny@mailbox.org
[2] University of Edinburgh, Edinburgh, UK

**Keywords:** Isotype · Vienna circle · Otto neurath · C. K. ogden
Basic English · Philosophy of language · Semiotics · International language

Isotype is a system of diagrammatic representation originally designed in the 1920s for conveying statistical information to visitors of the Vienna Museum of Society and Economy. The system was conceived by the museum's director, the Vienna Circle philosopher Otto Neurath (1882–1945), who saw it as an implementation of aspects of his philosophy of language and semiotic theory. In the 1930s Neurath came into contact with the English scholar C. K. Ogden (1889–1957), co-author with I. A. Richards (1893–1979) of *The Meaning of Meaning* [1], a classic text in semiotics, and inventor of the international language project Basic English. In the contact and resulting collaboration between Neurath and Ogden there is a clear convergence in their views. This abstract reports on recent research into Neurath, Ogden and their milieu, published principally in McElvenny [2] and McElvenny [3].

'Words divide – pictures unite' is Neurath's repeated slogan for Isotype [4–5]. It echoes his call for unified science: 'Metaphysical terms divide – scientific terms unite' [6]. In common with many philosophers of his time, Neurath believed that pictures, unlike words in language, possess an inherent connectedness to the world. According to Neurath, pictures show – and can only show – concrete, tangible objects; they are incapable of expressing the abstract entities that populate the metaphysician's world. The concreteness of pictures also makes them universally accessible: anyone should be able to understand a well-constructed pictorial diagram, regardless of their native language, level of education or cultural background.

Isotype was intended as an implementation of this philosophy of language, initially for the specific purpose of conveying statistical information. A typical example of Isotype in this mode can be seen in Fig. 1, which illustrates the different types of economies found in the world and their distribution over various population groups, with their sizes. The toothed wheel represents modern industrial economies, the hammer economies based around skilled trades and agriculture, and the bow and arrow hunter and gatherer economies with primitive agriculture. Each figure represents 100 million people: the outlined figures with hats represent Europeans; the brown figures with turbans represent 'orientals', Indians and Malays; the black figures Africans and 'mulattoes'; and the yellow figures with pointed hats represent 'Mongols'.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 800–802, 2018.
https://doi.org/10.1007/978-3-319-91376-6

**Fig. 1.** 'Economic systems of the earth' [7]

Ogden was among the contemporaries of Neurath who similarly accorded pictures high status for their transparency and faithfulness. The heart of Ogden's international language project Basic English [8] was a 'core vocabulary' of 850 words 'scientifically selected' for their reliability in picking out referents in the world. In choosing these 850 words, Ogden preferred nouns to verbs because – among other reasons – nouns generally name things that can be 'pictured', while verbs do not. The Basic core vocabulary was intended to provide the necessary building blocks that speakers of the language can use to paraphrase their concepts in simple terms. Paraphrases in terms of visual or otherwise observable properties were preferred over more elusive descriptions [9].

In the correspondence between Ogden and Neurath, beginning in the early 1930s and continuing until Neurath's death in 1945, we can see the two scholars discovering their common views on philosophy of language and semiotics and coming closer together on many points. In this period, Ogden, in his role of editor at Kegan Paul publishers, commissioned a number of books from Neurath on Isotype, *Isotype: international picture language* [5] and *Basic by Isotype* [10]. We see in the title of the first book Ogden's success in convincing Neurath to align his Isotype with the international language movement which, at the time, was a major interest in many quarters of academia and society at large. The second book used Isotype diagrams as a method to bootstrap the learning of Basic. Although Neurath was always cautious in pressing the claims of Isotype to being a complete linguistic system in itself, in the years of his collaboration with Ogden the system was used increasingly for instructions and narratives where a series of iconic pictures was presented stepping through an action or a sequence of events. Figure 2, for example, instructs parents to take their children to the doctor to be cured when they exhibit the symptoms of rickets.

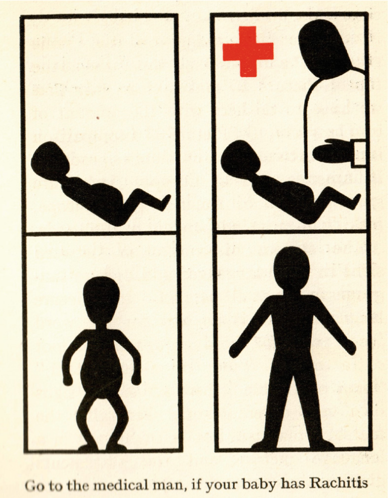


**Fig. 2.** Rachitis [5]

## References

1. Ogden, C.K., Richards, I.A.: The Meaning of Meaning. Kegan Paul, London (1923)
2. McElvenny, J.: International language and the everyday: contact and collaboration between C. K. Ogden, Rudolf Carnap and Otto Neurath. Br. J. Hist. Philos. **21**, 1193–1218 (2013). https://doi.org/10.1080/09608788.2013.825761
3. McElvenny, J.: Language and Meaning in the Age of Modernism: C. K. Ogden and His Contemporaries. University of Edinburgh Press, Edinburgh (2018)
4. Neurath, O.: Bildstatistik nach Wiener Methode in der Schule. In: Haller, R, Kinross, R., Neurath, O. (eds.) Gesammelte bildpädagogische Schriften, pp 265–336. Hölder-Pichler-Tempsky, Vienna (1991, 1933)
5. Neurath, O.: Isotype: International Picture Language. Kegan Paul, London (1936)
6. Neurath, O.: Unified science and psychology. In: Neurath, O. (ed.) Unified Science: the Vienna Circle Monograph Series Originally. English edition, Brian McGuinness (trans and ed), pp 1–23. D. Reidel, Dordrecht (1987, 1933)
7. Gesellschafts- und Wirtschaftsmuseum Wien Gesellschaft und Wirtschaft: bildstatistische Elementarwerk. Bibliographisches Institut, Leipzig (1930)
8. Ogden, C.K.: Basic English: International Second Language. (Edited by Graham, E.C.). Harcourt, Brace and World, New York (1968)
9. Ogden, C.K.: Work in progress (editorial). Psyche **10**, 1–30 (1929)
10. Neurath, O.: Basic by Isotype. Kegan Paul, London (1937)

# Diagrammatic Approaches in Computational Musicology: Some Theoretical and Philosophical Aspects

Moreno Andreatta[1,2], Carlos Agon[1], Alexandre Popoff[3],
and Andrée Ehresmann[4]

[1] CNRS-IRCAM-UPMC, Paris, France
{Moreno.Andreatta, Carlos.Agon}@ircam.fr
[2] IRMA/GREAM/USIAS, Université de Strasbourg, Strasbourg, France
[3] 119 Rue de Montreuil, Paris, France
al.popoff@free.fr
[4] LAMFA, Université de Picardie, Amiens, France
acehres@wanadoo.fr

**Abstract.** Despite a long historical relationship between mathematics and music, the use of diagrammatic approaches in computational musicology is a relatively recent phenomenon. Within the different branches of formal methods in music analysis, the so-called "transformational" paradigm has progressively shifted from an object-oriented to a graph-theoretical and categorical approach. Both graph theory and category theory make large use of diagrams which enable the description of the inner relationships of musical structures. In the categorical framework recently proposed by the authors, whose results are summarized and discussed in this abstract, musical transformations are viewed as natural transformations between chords represented as labelled graphs with vertices corresponding to the notes and arrows corresponding to musical transpositions and inversions operations. The diagrammatic approach also provides a very powerful conceptual tool that can have crucial theoretical implications for music cognition. We discuss this aspect by showing some deep connections between transformational music analysis and some mathematically-oriented directions in developmental psychology and cognition (such as Halford and Wilson's neostructuralistic approach, Ehresmann and Vanbremeersch's Memory Evolutive Systems, Phillips and Wilson's Categorical Compositionality, Fauconnier and Turner's Conceptual Blending and its structural extension proposed by Goguen) and epistemology (Gaston-Granger's "objectal" and "operational" duality).

**Keywords:** Graphs · Category theory · Transformational analysis
Computational musicology · Epistemology · Cognitive sciences

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 803–806, 2018.
https://doi.org/10.1007/978-3-319-91376-6

## 1 The Transformational Approach in Music Analysis

Within the different formal approaches in transformational music analysis, neo-Riemannian theory occupies a singular position by stressing a “dualistic” perspective on the Tone System based on inversional relations between major and minor chords. After a first algebraic formalization by David Lewin through the concept of *Generalized Interval System* [9], neo-Riemannian theory and analysis have developed more and more sophisticated tools by showing their deep categorical roots [5, 10].

According to this music-theoretical and analytical paradigm, there are three ways of transforming a major chord into a minor chord by preserving two common notes: the R transformation (as “relative”), that changes for example the C major chord into the A minor chord; the P transformation (as “parallel”) that changes the C major chord into the C minor chord; the L transformation (as “Leading-Tone operator”), changing the C major into the E minor chord. In a categorical framework, the neo-Riemannian operations are viewed as *natural transformations* between major and minor chords represented as labelled graphs with vertices corresponding to the notes and arrows corresponding to transposition and inversion operations [12−14]. By definition a transposition by $h$ semi-tons is an operation indicated by $T_h$ that sends a generic element $x$ of the cyclic group of order 12 (i.e. a pitch-class in the musical set-theoretical terminology) into $x + h$ (modulo 12). Similarly, one may define the generic inversion operation indicated by $I_k$ that sends a pitch-class $x$ into $k - x$ (always modulo 12). The diagrammatic approach to neo-Riemannian operations is shown in Fig. 1

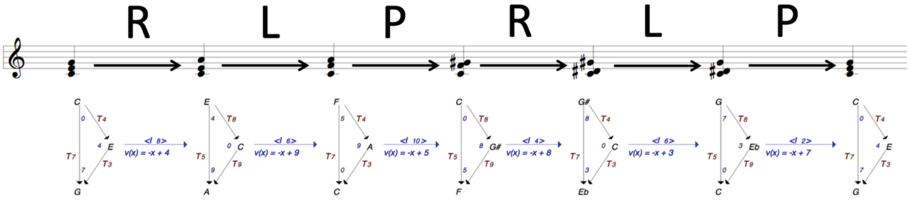


**Fig. 1.** The sequence of six neo-Riemannian transformations constituting the hexagonal cycle that generates the Tonnetz by musical transpositions. We have associated to the chord sequence the underlying diagrammatic framework displaying the transformations between major and minor chords viewed as labelled graphs with vertices corresponding to the notes and arrows corresponding to musical transposition. Although inversions $I_k$ do not play any role in this case, one may find so-called “negative isographies” $< I_k>$ acting on the arrows $T_k$ of the corresponding labelled graphs (for example, the first negative isography $<I_8>$ transforms $T_7$ into $T_5$, $T_4$ into $T_8$ and $T_3$ into $T_9$). Each negative isography is accompanied with the transformation acting on the vertices of the labelled graph (in the previous case, for example, the notes C, E and G are transformed respectively in the notes E, C and A via the map $-x + 4$).

## 2 Towards a Diagrammatic and Category-Based Music Cognition

Following some previous attempts at developing a category-based approach to creativity [2], the diagrammatic formalization of musical structures provides a very powerful conceptual tool that can have crucial theoretical implications for cognitive sciences and mathematical psychology. Combining the transformational approach, with a computational perspective has some crucial theoretical implications for cognitive sciences and mathematical psychology, as one may see by analyzing some mathematically-oriented directions in developmental psychology and cognition [3, 4, 6, 8, 11]. From an epistemological point of view, transformational analysis provides an instantiation, in the music domain, of Gilles-Gaston Granger's articulation between the "objectal" and the "operational" dimensions [7]. This duality was considered by the French epistemologist as the foundational basis for the very notion of "concept" in philosophy.

The use of diagrams within a categorical framework enables precisely to explain the shift from the object-oriented perspective provided by tradition musical *Set Theory* [1] to a "relational" approach in transformational music analysis [14].

## References

1. Andreatta, M., Bardez, J.-M., Rahn, J. (eds.): Around Set Theory. A French-American Musicological Meeting, IRCAM/Delatour France (2003)
2. Andreatta, M., Ehresmann, A., Guitart, R., Mazzola, G.: Towards a categorical theory of creativity for music, discourse, and cognition. In: Yust, J., Wild, J., Burgoyne, J.A. (eds.) MCM 2013. LNCS, vol. 7937, pp. 19–37. Springer, Heidelberg (2013)
3. Ehresmann, A.C., Vanbremeersch, J.-P.: Memory Evolutive Systems: Hierarchy, Emergence, Cognition. Elsevier, Amsterdam (2007)
4. Fauconnier, G., Turner, M.: The Way We Think: Conceptual Blending and the Mind's Hidden Complexities. New York, Basic Books (2002)
5. Fiore, T.M., Noll, T.: Commuting groups and the topos of triads. In: Agon, C., Andreatta, M., Assayag, G., Amiot, E., Bresson, J., Mandereau, J. (eds.) MCM 2011. LNCS (LNAI), vol. 6726, pp. 69–83. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21590-2_6
6. Goguen, J.: An introduction to algebraic semiotics, with application to user interface design. In: Nehaniv, C.L. (ed.) CMAA 1998. LNCS (LNAI), vol. 1562, pp. 242–291. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48834-0_15
7. Granger G.-G.: Formes, opérations, objets. Paris: Librairie Philosophique J. Vrin (1994)
8. Halford, G.S., Wilson, W.H.: A category-theory approach to cognitive development. Cogn. Psychol. **12**, 356–411 (1980)
9. Lewin, D.: Generalized Musical Intervals and Transformations. Yale University Press, New Haven (1987)
10. Mazzola, G., Andreatta, M.: From a categorical point of view: K-nets as limit denotators. Perspect. New Music **44**(2), 88–113 (2006)
11. Phillips, S., Wilson, W.H.: Categorical compositionality: a category theory explanation for the systematicity of human cognition. PLoS Computat. Biol. **6**(7), 1–14 (2010)

12. Popoff, A., Andreatta, M., Ehresmann, A.: A categorical generalization of klumpenhouwer networks. In: Collins, T., Meredith, D., Volk, A. (eds.) MCM 2015. LNCS (LNAI), vol. 9110, pp. 303–314. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20603-5_31
13. Popoff, A., Agon, C., Andreatta, M., Ehresmann, A.: From K-nets to PK-nets: a categorical approach. Perspect. New Music **54**(2), 5–63 (2017)
14. Popoff, A., Andreatta, M., Ehresmann, A.: Relational PK-nets for transformational music analysis. J. Math. Music (in press)

# Fruitful Over-Determination in Knot Diagrams

Silvia De Toffoli

Stanford University, Stanford, CA 94305, USA
`silviadt@stanford.edu`

**Abstract.** Thinking in mathematics is often mediated by diagrammatic notations which support specific manipulations. In this note, I address epistemic issues concerning knot diagrams. They form an effective notation, but they over-determine the target of study: knot types. Nevertheless, the over-determination contributes to their effectiveness because it underwrites the possibility of both exploiting topological intuition and performing automatic computations.

**Keywords:** Diagrams · Knot diagrams · Mathematical notations Over-determination

In the present abstract, I focus on the role of diagrams in knot theory. I consider a recent article: "More knots in knots: A study of classical knot diagrams" by Millett and Rich [4] to analyze two different ways in which mathematicians use knot diagrams: by visual manipulation and by computer assisted procedures.

Briefly, a *knot* is a simple (i.e. with no self-intersection) closed curve embedded in space. Generally, we are interested only in how a knot is knotted, and not in its specific shape, that is, we focus on *knot types*. In Fig. 1 are three diagrams of the trefoil knot. In order to present a knot, we can take a physical model or a picture of it, assign to it a name, or appeal to a smooth function. However, in practice it is common to use *knot diagrams*. These are essentially different from pictures of knots. Knot diagrams follow strict constraints which guarantee that all the relevant information is clearly displayed. Moreover, they do not only serve as illustrations, but they present a well-regimented operative dimension, which makes them effective in the practice of knot theory. Knot diagrams are usually considered as topological diagrams: their geometric shape is generally not relevant for the knot theorist. Accordingly, diagrams in Fig. 1(a) and (b) are equivalent. Moreover, we can define a series of operations on knot diagrams which allow us to connect all the different diagrams representing the same knot type: the three *Reidemeister Moves*. All three diagrams in Fig. 1 are equivalent modulo Reidemeister moves. When projecting knots on a surface, some information about the geometric shape of the knot is lost: infinitely many geometric knots give rise to the same knot diagram. Nevertheless, this should not be interpreted as a flaw of knot diagrams, but rather as an advantage. In fact, the lost information is irrelevant for the purposes of studying knot types.

Manders [conference talk] has stressed the problem of *over-determination* even in the case of knot diagrams. As we saw, the particular geometric shape

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 807–809, 2018.
https://doi.org/10.1007/978-3-319-91376-6

(a) (b) (c)

**Fig. 1.** Diagrams of the Trefoil Knot

of diagrams is not what interests the knot theorist, who will generally interpret a knot diagram topologically. Moreover, as explained in [2], we often study diagrams to track knot types, that is, modulo Reidemeister moves. Therefore, we can individuate at least two levels of 'extra information': geometric and topological.[1] Nevertheless, I claim that both ways in which a knot diagram over-determines a knot type is conducive to the practice of knot theory. This is because they respectively underwrite the possibility of: (1) Manipulating knot diagrams *via* "enhanced manipulative imagination;" and (2) 'Discretizing' knot diagrams in multiple ways.

One of the advantages of representing knots via knot diagrams is that the latter can be manipulated by exploiting cognitive abilities developed in our interaction with concrete objects. Given their geometric shape and their topological interpretation, we can easily 'visualize' many transformation that do not alter the corresponding knot, as in Fig. 1. In a previous work in collaboration with Giardino [2] we labeled the faculty at play "enhanced manipulative imagination" to refer to both its connection with manipulative abilities and the fact that it is enhanced by specific mathematical training. Brown [1, Ch. 6] also highlights the importance of knot diagrams' visual appearances. He claims that knot diagrams form an effective mathematical notation because, thanks to their visual features, we can calculate with them.

Therefore, whereas it is true that knot diagrams over-determine knot types, it is also the case that such representations trigger manipulative imagination exactly because they are geometric representations. In a recent article, Millett and Rich [4] aim at analyzing the compositional structure of knots and knot diagrams. Part of their analysis is done by visual inspection of the knot diagrams present in standard knot tables. This is possible because knot diagrams, although less redundant than knots, still present geometric information that allows human agents to exploit topological intuition.

After such visual inspection, Millett and Rich's results are derived from automatic computations that can be implemented after a discretization of knot diagrams. These can in fact be divided into discrete components canonically: the strands (curves going from one undercrossing to another undercrossing) and the crossings of which they are composed. As pointed out in [3, p. 152], this discretization underwrites the possibility of deploying algebraic and combinatorial

---

[1] That is, knot diagrams have a specific geometry (which distinguishes between diagrams in Fig. 1(a) and (b)), and a specific topology, (which distinguishes between diagrams in Fig. 1(a) and (c), but not between diagrams in Fig. 1(a) and (b)).

tools in the study of knots. We can get different discrete counterpart of knot diagrams: e.g. by (1) coding knot diagrams with numbers; (2) creating discrete graphs that can be interesting on their own or shed light on knot's properties; and (3) associating algebraic invariants to knots.

In the paper, I analyze all these discrete counterparts of knot diagrams and focus on their epistemic roles in the practice. For example, Millett and Rich use the *Dowker-Thistlethwaite* (DT) code. The DT code is a sequence of integers that can be easily obtained from a knot diagram by associating numbers to its crossings. Note that even if codes are used to study knots, they do not refer directly to knots, but only indirectly, *via* referring to knot diagrams. Because of their intermediary status, knot diagrams play a special epistemic role in the practice knot theory.

Conceiving mathematics as a human activity and not only as a timeless body of truths leads to acknowledging the importance of mathematicians' daily practices. Mathematicians' reasoning is mediated by external material artifacts, which support manipulations corresponding to mathematical operations. In knot theory, the importance of diagrammatic notations is evident. Knot diagrams are both interesting for their own mathematical properties and used to refer to (mathematical) knots. Knot diagrams are an example of a perspicuous and transparent notation because they can be easily interpreted and manipulated correctly. They allow mathematicians to exploit cognitive abilities they developed from manipulating concrete objects by re-deploying them in the abstract domain of mathematics. In this respect, they differ from their code counterparts. However, by their own 'discretizable' nature, knot diagram bridge the domain of low-dimensional topology proper of knots to the one of discrete mathematics. The possibility of associating canonically discrete counterparts to knot diagrams underwrites the possibility of extracting their combinatorial properties and encoding them into sequences of numbers. Such encodings are useful in the practice because they can be used to perform automatic computations.

Both methods of inquiry mediated by knot diagrams, visual analysis and automatic computations, are essential in the practice of contemporary knot theory.

## References

1. Brown, J.R.: Philosophy of Mathematics: A Contemporary Introduction to the World of Proofs and Pictures. Routledge (2008)
2. De Toffoli, S., Giardino, V.: Roles and forms of diagrams in knot theory. Erkenntnis **79**(3), 829–842 (2014)
3. Ippoliti, E.: Ways of advancing knowledge. a lesson from knot theory and topology. In: Ippoliti, E., Sterpetti, F., Nickles, T. (eds.) Models and Inferences in Science. SAPERE, vol. 25, pp. 147–172. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-28163-6_9
4. Millett K.C., Rich, A.: More knots in knots: a study of classical knot diagrams. J. Knot Theory Ramif. 1750046 (2017)

# The Epistemology of Mathematical Necessity

Catherine Legg

Deakin University, Burwood, Australia
c.legg@deakin.edu.au

**Abstract.** It seems possible to know that a mathematical claim is necessarily true by inspecting a diagrammatic proof. Yet how does this work, given that human perception seems to just (as Hume assumed) 'show us particular objects in front of us'? I draw on Peirce's account of perception to answer this question. Peirce considered mathematics as experimental a science as physics. Drawing on an example, I highlight the existence of a primitive constraint or blocking function in our thinking which we might call 'the hardness of the mathematical must'.

**Keywords:** Necessity · Epistemology · Mathematics · Hume · Peirce

## 1 Introduction

We can come to know that a mathematical claim is true by inspecting a diagram, as in Fig. 1. What is being perceived here? Not just that $2 \times 3$ is $3 \times 2$, but that $2 \times 3$ must be $3 \times 2$. It is clear that trying to create an option such as $2 \times 3 = 3 \times 3$ would be futile. We could of course prove the same claim in a more stepwise, symbolic manner, but this diagram seems to give us everything we need to ascertain the necessary truth.

Yet how exactly does this work? Much mainstream analytic epistemology arguably makes such knowledge-gathering seem impossible [1], through a materialist understanding of perception, deriving ultimately from Hume's empiricism, according to which, we might say, 'experience only shows us the particular objects in front of us'. This produces a highly sceptical treatment of modality, which Hume famously used to undermine so-called causal necessity [1]. Attempts to challenge this often meet intimidating charges of anti-naturalism.

## 2 Hume's Legacy and Its Challenges

Hume's particular empiricism led him to coin a supposedly common-sense maxim, widely taken for granted today: ***"There are no necessary connections between distinct existences".*** The maxim follows from two key Humean claims about perception: (i) it is *passive*—involving nothing more than 'registering' the impact of individual physical objects on the sense organs, (ii) it is *atomistic*—ideas are only distinct if fully separable in imagination. Consequently Hume banishes from his epistemology *abstract ideas*, whose determinable properties lack determination (e.g. a 'general triangle'). Allowing these would render the mind active in choosing which determinables to elide. The early

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 810–813, 2018.
https://doi.org/10.1007/978-3-319-91376-6

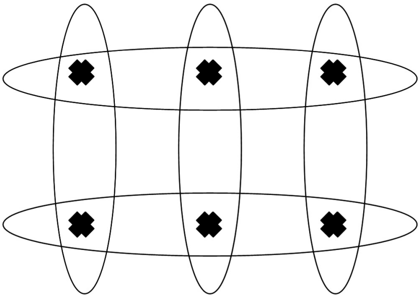


**Fig. 1.** Why 2 × 3 = 3 × 2

modern Hume sees this denial as properly naturalistic against scholasticism. Consider Fig. 1. Does it display a necessary connection between distinct existences? Well, what are the 'existences'? Humeans have a number of choices here:

(i) *Physical Marks:* The relevant existences are 5 oval-shaped and 6 star-shaped marks, exactly as inscribed on the page. This choice appears natural given the organisation of our visual field and—invoking Hume's separate imaginability criterion—we can imagine each shape existing alone on the page. But there *are* necessary connections between these objects in Fig. 1. For instance one cannot remove stars from the vertical ovals without changing the number of stars in the horizontal ovals. Interpreted thus, then, Hume's maxim seems simply incorrect.

(ii) *Abstract Objects:* Alternatively one might claim that Fig. 1 displays a truth about something more ideal—e.g. three '2s' and two '3s'. These are arguably not distinct, since 2 is made up of 'two ones' and 3 of 'three ones', so 2 is a proper part of 3. At this point, then, Hume might defend his maxim by stating that Fig. 1 expresses relations between ideas, not matters of fact, and only the latter is accessed through perception, whilst the former is mere semantic stipulation. But this is somewhat unsatisfying. If mathematics merely concerns a world of ideas, how are physical diagrams so startlingly effective? Furthermore, now Hume's maxim seems to beg the question. Is he not arbitrarily ruling out that we perceive existences between which necessary connections hold, effectively stating: "There are no necessary connections between distinct existences, which are those existences without necessary connections".

(iii) *"Both":* One might consider combining the two views as follows: the existences are ovals and stars *and* '2s' and '3s'. However this raises tricky questions about the relationship between the physical marks and the numerical objects. If they are all separate, why include physical marks in the diagram? Why not 'cut to the chase' and just include the numerical objects? Obviously that is impossible, which points to our preferred interpretation.

(iv) *"Hybrid…but not both":* Attribute to the physical marks and numbers *partial identities.* What does this mean? Just that 'twoness' is a property which may be abstracted from two star-shaped marks on the page, while precisely not being separable from them. Abstraction without separation is essential for all *structural reasoning*. This includes our example since it turns on recognising that two abstractable but not separable structures: 2 × 3 (two rows of three stars) and 3 × 2 (three columns of two stars) – are in fact identical. Structural reasoning is surely a significant part of mathematics. So what theory of perception could do justice to it?

## 3 A Peircean Approach

Peirce suggests we need to give separate, interlocking, accounts of: (i) immediate experience of objects: the *percept*, (ii) the truth of symbols derived from that experience: the *perceptual judgment*. The percept is a non-cognitive direct encounter with some object. It is not a Humean idea, nor does it express truth-claims, it "simply knocks at the portal of my soul and stands there in the doorway" [5]. On the other hand, the perceptual judgement takes propositional (subject-predicate) form, and its interpretation opens to the community of inquiry in an endless series of judgments, each member of which is logically related to prior members. The perceptual judgement does not copy the percept, as they are too unlike one another. How do they relate? Contra Hume, percepts *cause perceptual judgements, while not being the source of their content*. Human evolution ensures that each percept causes "direct and uncontrollable interpretations". This process can and must be trained through cultivating appropriate mental *habits*, through public criticism in a common language.

So what kind of 'existence' is the mathematical percept? We should take seriously Peirce's repeated claims that mathematics is as experimental a science as physics, but the mathematician's laboratory is the diagram [5]. Again consider Fig. 1. First note that the mathematical percept, like all percepts, is strictly impossible to describe in words. However, when I 'got' this proof, I suddenly grasped the horizontal and vertical star-arrangements as one, as if the same 5 ovals were 'holding both together'. 'Holding together' is a metaphor, as the arrangements are not strictly *parts* of the diagram (as not separable, only abstractable). But looking at the diagram and thinking about abstracting other arrangements from it (such as three threes), I could feel myself not being able to think of them. We might call this primitive blocking or constraint, in homage to Wittgenstein, 'the hardness of the mathematical must'. With my prior mathematical training, this prompts me to an 'uncontrollable interpretation' that the proposition 2 × 3 = 3 × 2 must be true. No matter how hard I try to interpret Fig. 1 as 2 × 3 = 3 × 3 – I simply cannot think that way. *(Try it yourself…)* Thus Peirce notes that although mathematics deals with a world of ideas, not material objects, its discoveries are something to which our minds are forced. My felt 'hardness' now becomes the necessity of the perceptual judgment: 2 × 3 = 3 × 2.

Thus, despite many philosophers' bafflement, we do perceive necessity. I have argued that perception is in fact the *only* way we come to know necessity, as all necessary reasoning involves experimenting on diagrams to determine structural dependencies [2–4]. Necessary truths are abstract able from physical marks and this is not an ontological reification but an epistemic capacity. Arguably the whole concept of 'abstract object', so problematic in recent philosophy of mathematics, arises from not understanding that (contra Hume) one can abstract without separating.

## References

1. Hume, D.: A Treatise of Human Nature (Edited by Selby-Bigge). Oxford (1739-40/1978)
2. Legg, C., Franklin, J.: Perceiving necessity. Pac. Philos. Q. **98**(3), 320–343 (2017)
3. Legg, C.: What is a logical diagram? In: Moktefi, A., Shin, S.-J. (eds.) Visual Reasoning with Diagrams, pp. 1–18. Springer, Basel (2013)
4. Legg, C.: The hardness of the iconic must: can Peirce's existential graphs assist modal epistemology? Philos. Math. **20**(1), 1–24 (2012)
5. Peirce, C.S.: Collected Papers (Edited by Hartshorne, A., Weiss, P., Burks, A.W.). Harvard (1931–1958)

# The Central Role of Diagrams in Algebraic Topology

Emily R. Grosholz

The Pennsylvania State University, University Park, PA, USA
erg2@psu.edu

**Abstract.** The great achievement of Cantor and Dedekind in the 19th c. was not the arithmetization of the geometrical line, but rather the use of the notion of set to extend the number system to the reals and the transfinite numbers, and to recast the line as a set of points and numbers systems as sets of numbers: this prepares the ground for the emergence of topology and then algebraic topology in the 20th c. and the novel diagrams that accompany its presentation in the textbook written by Singer and Thorpe.

**Keywords:** Real number · Algebraic topology · Homology · Manifold

## 1 Introduction

Set Theory and Group Theory in the 19th century play an important role in the development of Topology and then Algebraic Topology in the 20th century, and novel diagrams play a central role in this development. The axioms defining a topological space were first formulated successfully by Felix Hausdorff. Problems in analysis (such as the search for new, more abstract definitions of function and integral), and the development of Set Theory by Cantor, had led mathematicians like René Fréchet to search for ways of talking about abstract sets and spaces, whose components were not points or real numbers, but simply elements (Manheim 1964, pp. 116–119). It was Hausdorff who first noted the distinction between the traditional understanding of distance and limit, and the notion of neighborhood. The concept of distance is tied to a metric, and that of a limit depends on a relation of countability, so Hausdorff selected the notion of neighborhood as more fundamental. He also "knew how to choose, among the axioms of Hilbert on neighborhoods in the plane, those which were able to give his theory both the desirable precision and generality." (Manheim 1964, pp. 122–123). Hausdorff spaces are spaces where there are enough open sets for any two distinct elements to be contained in two open sets that do not intersect.

## 2 Homotopy and Homology Theory

First, we explore the role of diagrams in homotopy theory. The key to Algebraic Topology is to construe, or construct, a topological space as a group. What then should the elements be? The points of a topological space itself, or even open subsets, would be unwieldy as group elements because there are just too many of them. The elements

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 814–816, 2018.
https://doi.org/10.1007/978-3-319-91376-6

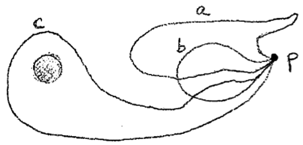


**Fig. 1.** Homotopy Equivalence Classes, on Plane with Hole

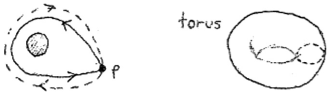


**Fig. 2.** Homotopy Equivalence Classes: Plane with a Hole, and Torus

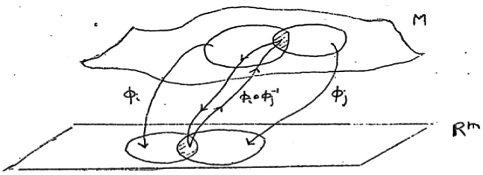


**Fig. 3.** A Smooth Manifold: Hausdorff Space and Collection of Maps

must, however, be something is it possible to find in every Hausdorff space. One choice is functions called paths, which map the unit interval {[0, 1] onto the topological space. Requiring that 0 and 1 get mapped to the same point *p* of the space, and that *p* is the same for all paths, it is possible to cover the whole space with these 'loops.' A binary operation of two path-functions *a* and *b* can be defined by saying that *ab* maps [0,1] to the space (i.e. *ab* is itself a path-function) when *ab*(0) = *a*(0), *ab*(1) = a(1) = *b*(0), and *ab*(1) = b(1). That is, *ab* maps the unit interval around two loops, images of *a* and *b*.

However, this set of elements has a problem: there are too many of them. So far, we only succeeded in assigning a group with an infinite number of elements to every topological space. To make the association useful, we define as elements not paths, but equivalence classes of paths, via the equivalence relation homotopy. A homotopy exists between functions (that is, two functions are homotopic) when they have the same domain, and one function can be continuously deformed into another. In formal terms, this condition is expressed $a \simeq b$ (*a* is homotopic to *b*) when there exists a continuous map *F* such that F:[0,1] x [0,1] $\rightarrow$ T (the topological space) and F(*x*,0) = a (*x*) and F(*x*,1) = b(*x*). Consider the situation when the topological space (say the Euclidean plane) has a hole in it. The 'loops' *a* and *b* can be continuously deformed, shrunk back, until they collapse in the point *p*.

This means that they are in the same homotopy equivalence class as the identity path-function which maps all of [0,1] to *p*. But there is no way to draw 'loop' *c* back to *p* unless it breaks to get around the hole; but then the deformation is not continuous. Then *c* must be in a homotopy equivalence class different from the one which includes *a*, *b*, and the identity (Singer and Thorpe 1967, Chap. 3). This construction of a homotopy group, whose elements are homotopy equivalence classes, offers an algebraic structure which distinguishes between topological spaces which exhibit significantly different features. For instance, the Euclidean plane has no holes, and so its (so called) fundamental group structure has only one equivalence class, the identity, because all the loops can be collapsed into the single point $p$. The same is true of the surface of a sphere. However, a plane with a hole in it has a fundamental group which is isomorphic to the integers.

Once around the hole corresponds to 1, twice to 2 and so forth; once around the hole the other way corresponds to −1 etc. and the identity corresponds to 0. The torus has two holes, one in the center and one inside, so its fundamental group corresponds to the Cartesian product of the set of integers with itself, the pairs $(x,y)$, with $x$ and $y$ integers. There are many other ways to develop this thought, for instance, mapping sphere-like images onto the space instead of loops, or changing the notion of equivalence from homotopy to homology.

Manifolds are good focus for homology Theory. Although they come in complicated shapes and curves, their definition insures that they are locally homeomorphic to $n$-dimensional Euclidean space (the prototype of any $n$-dimensional vector space) (Singer and Thorpe 1967, Chap. 5). That is, for every point $x$ of the manifold there is an invertible function $\phi$ that maps an open set (a neighborhood) around $x$ one-to-one and onto an open set in $\mathbf{R}^m$. We will confine ourselves to smooth manifolds, where functions behave nicely on the overlap of neighborhoods. Formally, a smooth manifold is a pair $(M, \phi)$ where $M$ is a Hausdorff topological space and $\phi$ is a collection of maps such that the domain of the $\phi$ are small open neighborhoods of $M$ which completely cover $M$, and each one maps these neighborhoods homeomorphically onto neighborhoods in $\mathbf{R}^m$, so that the coordinate functions $(r_1, r_2, \ldots r_m)$ in $\mathbf{R}^m$ correspond to local coordinate functions on the manifold $(m_1, m_2, \ldots m_m)$ via the local representation $m_i = r_i \circ \phi$. In addition, $\phi_i \circ \phi_j{}^{-1}$ must be a smooth map on the overlap of their respective domains, going from the image of $\phi_i$ in $\mathbf{R}^m$ back to $\mathbf{R}^m$.

This set $\Phi$ of homeomorphisms insures us that, at least locally, whatever structure we can find in $\mathbf{R}^m$ (a vector space is a likely place to look for structure) will have an analogue on the manifold, and further than that, whatever structure we get locally we can translate smoothly from neighborhood to neighborhood, and so understand the global structure. This search for structure on the manifold, translating back and forth between the manifold and $\mathbf{R}^m$, will lead us through the algebraic structure of vector spaces, linear transformations, Grassman algebra and finally back to a glimpse of De Rham's Theorem.

**Acknowledgments.** This abstract draws on Part IV, Chapter 2 of my book *Great Circles*, which is forthcoming in 2018 from Springer.

## References

1. Manheim, J.H.: The Genesis of Point Set Topology. Pergamon Press, New York (1964)
2. Singer, I.M., Thorpe, J.A.: Lecture Notes on Elementary Topology and Geometry. Foresman and Company, Glenview (1967)

# Can Spontaneous Diagram Use be Promoted in Math Word Problem Solving?

Hiroaki Ayabe(✉) and Emmanuel Manalo

Graduate School of Education, Kyoto University, Kyoto 606-8501, Japan
`ayabe.hiroaki.56c@kyoto-u.jp`,
`manalo.emmanuel.3z@kyoto-u.ac.jp`

**Abstract.** This study investigated the efficacy of providing instruction and practice to promote spontaneous diagram use in the math word problem solving of 70 students (mean age = 14 years) participating in a quasi-experiment in a real school setting. The experiment required the students to solve math word problems which could be more efficaciously solved by using appropriate diagrams (based on the structure of the problems). In the pretest, diagrams were rarely used by the students. However, following an intervention (provided as instruction and practice in construction and use of the appropriate diagrams), the students' spontaneity in diagram use and their correct answer rates significantly increased. The intervention encouraged the use of specific diagrams to match features of the problem to solve. This demonstrated the domain specificity of diagram use in math word problem solving. Furthermore, the level of diagram use and correct answer rates maintained over the experimental period. We also confirmed that cognitive load (based on student self reports of cognitive effort) decreased as a consequence of the intervention provided, thereby providing one explanation for the increased spontaneity we observed.

**Keywords:** Spontaneous diagram production · Math problem solving
Strategy use advice and encouragement · Diagram use instruction
Skills practice

## 1 Introduction

It is necessary for us to use mathematics in various facets of everyday life. Developing skills in solving math word problems are particularly important because such problems contextualize the need to apply math knowledge and skills, rendering them as useful practice for real life situations. However, the correct answer rates tend to be low and, in general, students do not use diagrams spontaneously when solving such problems even though diagrams are considered effective tools for solving them [1].

Previous studies suggest that instructions about diagrams are effective in encouraging students to construct the appropriate diagrams, but such instruction did not always improve the correct answer rates [2]. But, in a recent study in the context of communication, students' spontaneous diagrams use was increased and maintained by using an intervention that added practice to instruction [3]. This suggested that practice lowered cognitive cost associated with diagram production. Therefore, in the present

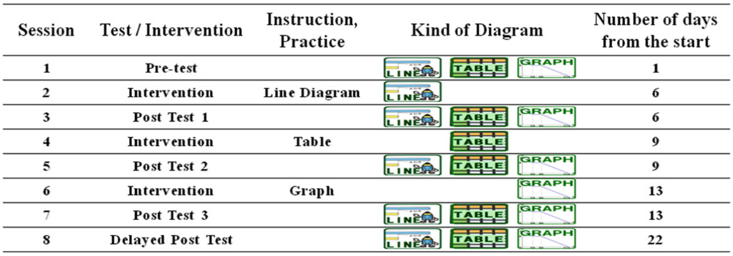
© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 817–820, 2018.
https://doi.org/10.1007/978-3-319-91376-6

study, we hypothesized that an intervention combining instructions and practice would likewise promote spontaneous diagram use by reducing cognitive cost, and that such use would also result in improving the correct answer rates in math word problem solving. One of our aims was to elucidate the mechanism of spontaneous diagram use by considering cognitive load [4] as factor influencing such use.

## 2 Method

Participants were 70 8th-grade students (female = 37; mean age = 14 years). We employed a pre- and post-test design, with intervention phases (baseline, interventions on diagram knowledge, delayed posttest), and with task factors (line diagram, table, graph). Interventions of instruction and practice (3 separate sessions) and tests (5 times) were carried out over 5 days in an actual classroom setting, with each session being 40 min in duration (see Table 1).

**Table 1.** Implementation plan



| Session | Test / Intervention | Instruction, Practice | Kind of Diagram | Number of days from the start |
|---|---|---|---|---|
| 1 | Pre-test | | LINE<br>TABLE<br>GRAPH | 1 |
| 2 | Intervention | Line Diagram | LINE | 6 |
| 3 | Post Test 1 | | LINE<br>TABLE<br>GRAPH | 6 |
| 4 | Intervention | Table | TABLE | 9 |
| 5 | Post Test 2 | | LINE<br>TABLE<br>GRAPH | 9 |
| 6 | Intervention | Graph | GRAPH | 13 |
| 7 | Post Test 3 | | LINE<br>TABLE<br>GRAPH | 13 |
| 8 | Delayed Post Test | | LINE<br>TABLE<br>GRAPH | 22 |

A questionnaire was administered immediately after each task. In the intervention sessions, instructions to develop semantic knowledge and procedural knowledge for use of each kind of diagram (line diagram, table, and graph) were provided. Equivalent math word problems (isomorphic in structure and requirements), in which each of these kinds of diagrams were required to more efficiently arrive at their solutions, were created and used in the instructions and tests (different problems were used in instructions and tests). All the math problems were presented in sentences only and did not include expressions to explicitly induce the use of diagrams.

Eight minutes were allowed for solving each problem, and no feedback was provided. The questionnaire (using 10-point Likert-type scales) measured intrinsic cognitive load (load imposed by the given task; 4 items), germane cognitive load (load that can be allocated for solving the task; 4 items), and expectancy-value perceptions (10 items). The diagrams included on the students' answer sheets were scored based on rubrics that allocated 0–2 points depending on appropriateness and detail.

## 3 Results and Discussion

Comparisons of scores at baseline and delayed posttest phases revealed that, in all diagram problem types, cognitive load *decreased*, and diagram scores and correct answer rates *increased* ($p < .01$ in all cases). Judging from the changes in scores that occurred between the instructions, each intervention affected only one problem type – the one that it specifically targeted (e.g., table-type problems only). These effects of the interventions were *maintained* even in the delayed posttests (see Fig. 1).

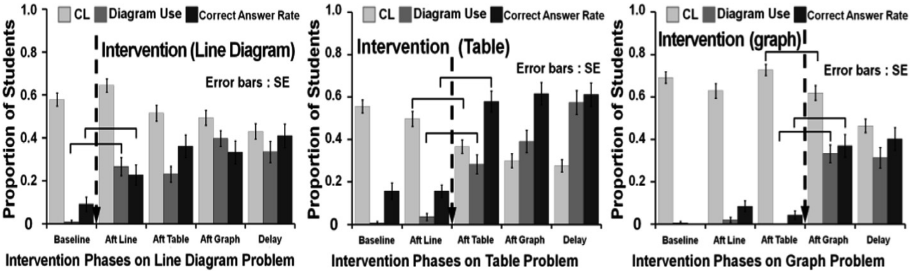


**Fig. 1.** Mean proportions of intrinsic cognitive load (CL), diagram use, and correct answer rates for each of the problem types (line diagram, table, and graph) across the phases of the study

Germane cognitive load also increased in all problem types ($p < .01$ in all cases). Significant correlations were also found between germane cognitive load and participant scores on the expectancy-value scales ($r = .45 - .55$).

The results of the present study demonstrate that interventions combining instruction with opportunities for practice successfully facilitated students' spontaneous diagram use and improved their correct answer rates. Thus, the hypothesis we posed was supported. The results suggest that intrinsic cognitive load in diagram production can be reduced by cultivating the necessary knowledge and skills in construction. The freed up cognitive resources can then be utilized as germane load, which can be deployed for actual solving/cognitive processing of the problems given. Furthermore, the relationship between germane cognitive load and motivation (expectancy and value) scores may provide an explanation for the maintenance of diagram use (i.e., higher expectation of success in being able to use diagrams + higher perception of value of using diagrams = higher motivation to use). The findings of this study also indicate the domain specificity of the diagram types and associated instructions.

## References

1. Uesaka, Y., Manalo, E., Ichikawa, S.: What kinds of perceptions and daily learning behaviors promote students' use of diagrams in mathematics problem solving? Learn. Instr. **17**, 322–335 (2007)
2. Uesaka, Y., Manalo, E.: Peer instruction as a way of promoting spontaneous use of diagrams when solving math word problems. In: Proceedings of the 29th Annual Cognitive Science Society, pp. 677–682. Cognitive Science Society, Austin, Texas (2007)
3. Manalo, E., Uesaka, Y.: Hint, instruction, and practice: the necessary components for promoting spontaneous diagram use in students' written work? In: Jamnik, M., Uesaka, Y., Elzer Schwartz, S. (eds.) Diagrams 2016. LNCS, vol. 9781, pp. 157–171. Springer, Cham (2016)
4. Paas, F., Sweller, J.: An evolutionary upgrade of cognitive load theory: using the human motor system and collaboration to support the learning of complex cognitive tasks. Educ. Psychol. Rev. **24**, 27–45 (2011)

# Cognitive Control as an Underpinning of Relational Reasoning from Diagrams

Leonidas A. A. Doumas[1], Robert G. Morrison[2],
and Lindsey Engle Richland[3]

[1] University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
alex.doumas@ed.ac.uk
[2] Loyola University Chicago, 1032 W Sheridan Rd, Chicago, IL 60660, USA
rmorrison@luc.edu
[3] University of Chicago, 1126 E. 59th Street, Chicago, IL 60637, USA
lrichland@uchicago.edu

**Abstract.** Diagrams are powerful opportunities for attending to abstract relationships, yet this is cognitively effortful, and individual differences in how reasoners benefit from such opportunities are not well understood. We posit that cognitive control is critical for both abstracting relational representations and using these representations analogically in subsequent problem solving. This hypothesis is evaluated using simulations in a symbolic connectionist model of analogical thinking, DORA/LISA (Discovery Of Relations by Analogy; Doumas, Hummel, & Sandhofer, 2008). Varying the base level of lateral inhibition in DORA affected the ability to learn and draw inferences from relational representations. These simulated accuracy rates and errors observed in children learning from repeated experiences in reasoning from geometric diagrams.

**Keywords:** Analogy · Diagrams · Cognitive control · Reasoning

## 1 Introduction

### 1.1 Empirical Trajectories of Children Learning from Geometric Diagrams

Diagrams are powerful opportunities for grappling with and learning abstract relationships, for example learning the relations between elements in an ecosystem rather than simply memorizing the objects within the system. Further, what is crucial from any diagrammatic learning opportunity is the ability to use this relational knowledge in a new context or with new materials, beyond simply understanding the initial presentation. This is cognitively effortful, however, and individual differences in how reasoners benefit from such relational learning opportunities are not well understood. We describe a computational simulation that examines how cognitive control of attention enables relational learning from visual stimuli such as diagrams. Specifically, we propose that cognitive control is critical for both abstracting relational representations from that visual stimuli, and to the ability to use these representations in new problem solving.

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 821–824, 2018.
https://doi.org/10.1007/978-3-319-91376-6

This study draws on extant longitudinal data from children who viewed and solved geometric analogy problems repeatedly over six months [1]. Each set of stimuli were geometric diagrams that contained a key set of relationships, such as *inside* (see Fig. 1). Children were presented with A:B :: C:D problems in which they had to draw the D term to make a valid analogy. Children's performance could be categorized into three distinct learning trajectories: analogical reasoners throughout, non-analogical reasoners throughout, and transitional - those who start non-analogical and grew to be analogical.

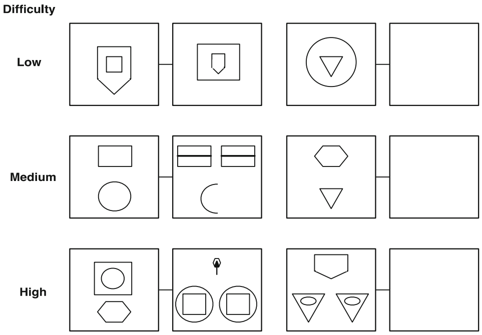


**Fig. 1.** Analogy problems varying in complexity based on those in Hosenfeld et al. (1997) [1].

Developments in children's analogical reasoning are traditionally attributed either to increased working memory resources due to maturation [2] or accretion of knowledge relevant to the particular task [3]. Both principles have empirical support, so in order to develop a comprehensive framework for how knowledge accretion and individual differences in cognitive skills together affect learning from diagrams, we test a theory for their integration using computational simulations in a symbolic connectionist model of analogical thinking, DORA/LISA (Discovery Of Relations by Analogy) [4].

### 1.2 Model Description

LISA [5] is a symbolic-connectionist model of analogy and relational reasoning. DORA [4] is a model, based on LISA, that learns structured (i.e., symbolic) representations of properties and relations from unstructured inputs. That is, DORA provides an account of how the structured relational representations LISA uses in the service of relational reasoning can be learned from examples. These models account for over 40 phenomena from the literature on children's and adults' relation learning [see 4–7].

## 2 Computational Simulations of Learning Trajectories

We hypothesized that differences between the children's analogical diagram learning trajectories [1] were at least partially a product of differences in working memory. We simulated these differences in LISA/DORA by varying levels of lateral inhibition. In LISA, inhibition is critical to the selection of information for processing in working memory. Inhibition determines LISA's intrinsically limited working-memory capacity [4], controls its ability to select items for placement into working memory and also regulates its ability to control the spreading of activation in the recipient.

We defined three groups following the behavioral data: (1) non-analogical, (2) transitional, and (3) analogical. We ran 100 simulations for each group. During each simulation we chose an inhibition level from a normal distribution with a mean of .4 for the non-analogical group, .6 for the transitional group, and .8 for the analogical group (each distribution had a SD = .2). For each simulation we ran 800 learning trials and checked the quality of the representations DORA had learned during the last 100 trials after each 100 trials. Quality was calculated as the mean of connection weights to relevant features (i.e., those defining a specific transformation or role of a transformation) divided by the mean of all other connection weights + 1. A higher quality denoted stronger connections to the semantics defining a specific transformation relative to all other connections (i.e., a more pristine representation of the transformation).

As observed in Fig. 2, DORA/LISA's growth trajectories closely followed the behavioral data. Like the non-analogical children, LISA/DORA with a low inhibition level performed poorly throughout. Like the transitional children, LISA/DORA with a medium inhibition level started slow but improved slowly. Like the analogical children, LISA/DORA with a high inhibition level performed well virtually from the start and maintained this throughout.

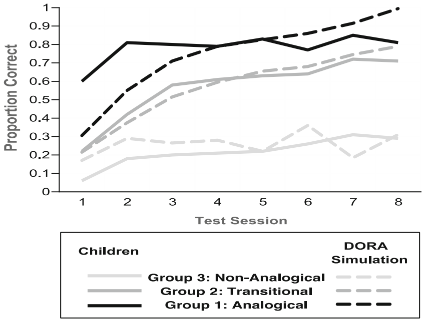


**Fig. 2.** Results from Hosenfeld et al., (1997 [1]) and LISA simulations. Simulation results were the result of training in DORA followed by reasoning in LISA. Groups were created solely by changing DORA/LISA's working-memory capacity (i.e., adjusting lateral inhibition).

## 3 Summary

In sum, children's learning from diagrams may vary due to their level of internal inhibitory control over the noise abstracted from diagrams as inputs, which may impact how likely they are to build and use relational representations from diagrams.

## References

1. Hosenfeld, B., van der Maas, H.L.J., van den Boom, D.: Indicators of discontinuous change in the development of analogical reasoning. J. Exp. Child Psychol. **64**, 367–395 (1997)
2. Richland, L.E., Morrison, R.G., Holyoak, K.J.: Children's development of analogical reasoning: Insights from scene analogy problems. J. Exp. Child Psychol. **94**, 249–273 (2006)
3. Rattermann, M.J., Gentner, D.: More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. Cogn. Dev. **13**, 453–478 (1998)
4. Doumas, L.A.A., Hummel, J.E., Sandhofer, C.M.: A theory of the discovery and predication of relational concepts. Psychol. Rev. **115**, 1–43 (2008)
5. Hummel, J.E., Holyoak, K.J.: A symbolic-connectionist theory of relational inference and generalization. Psychol. Rev. **110**, 220–264 (2003)
6. Morrison, R.G., et al.: A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. J. Cogn. Neurosci. **16**, 260–271 (2004)
7. Morrison, R.G., Doumas, L.L., Richland, L.E.: A computational account of children's analogical reasoning: balancing inhibitory control in working memory and relational representation. Dev. Sci. **14**(3), 516–529 (2011)

# Intersemiotic Translation: Transcreation and Diagrams

Letícia Vitra[1] and João Queiroz[2]

[1] Linnaeus University, Växjö, Sweden
leticiaavitral@gmail.com
[2] Federal University of Juiz de Fora, Juiz de Fora, Brazil
queirozj@gmail.com

**Abstract.** Creative translation as a method for investigation was an idea systematically explored by the Brazilian poet and translator Haroldo de Campos. According to Campos' approach, creative translation corresponds to the transcreation of a multi-level system of constraints, "selected" and revealed by the target system. We intend to describe this process as diagrammatic (sensu Peirce), in which the physicality of its source and target systems have the ontology of a relation. Hence our approach is a tentative association of Jakobson's concept of intersemiotic translation, with De Campos' notion of transcreation based on Peirce's notion of diagrams. We intend to describe it by taking the following arguments into consideration: I. Intersemiotic translation can be described as fundamentally triadic phenomenon, that involves the selection and interpretation of properties and methods from one semiotic system to be translated into another semiotic system, bearing the production of an interpretative effect in the latter, that is analogous to the interpretative effect produced by the former. II. Intersemiotic translation is a method of investigation. As a mainly iconic process, it produces a sign that signifies by means of its own qualities and structures: this is a well-known property of iconic signs, namely operational criterion of icons.

## 1 What Is the Role of Intersemiotic Translation for Discovery?

The idea of creative translation as a method for literary investigation was systematically explored by the Brazilian poet and translator Haroldo de Campos. According to Campos' approach, mainly based on Roman Jakobson [1] and Walter Benjamin [2], creative translation corresponds to the *transcreation* of a multi-level system of constraints, "selected" and revealed by the target-sign, an idea strongly inspired on Peirce's mature notion of iconicity. For Campos, *transcreation* is an iconic operation on the physicality of semiosis. Here we generalize this notion to intersemiotic translation phenomena. We associate Jakobson's concept of intersemiotic translation with De Campos' notion of transcreation. This association is based on Peirce's concept of diagrams as icons of relation. As a mainly iconic process, intersemiotic translation produces a sign that signifies by means of its own qualities and relational structures: this is a well-known property of iconic signs, namely *operational criterion of icons* (see Sect. 2.) [3]. In order

© Springer International Publishing AG, part of Springer Nature 2018
P. Chapman et al. (Eds.): Diagrams 2018, LNAI 10871, pp. 825–828, 2018.
https://doi.org/10.1007/978-3-319-91376-6

to properly investigate processes of intersemiotic translation based on those arguments, we are going to analyze a case of translation from architecture (Fig. 1(a)) to photography: the exhibitions named *Fachwerkhäuser des Siegener Industriesgebietes,* by the German photographers Bernd and Hilla Becher (Fig. 1(b)).


(a) (b)

**Fig. 1.** The architectural landscape of the Siegerland region (the source of the intersemiotic translation) and an example of a photographic grid from the *Fachwerkhäuser des Siegener Industriesgebietes* exhibition (the target-sign of the intersemiotic translation). (Sources: http://www.freudenberg-stadt.de/ and http://westfalium.de/2015/10/30/mettingen-die-kunst-des-aufbewahrens/)

## 2 Intersemiotic Translation as Operational Icon

The icon is operationally defined [3] as a sign whose manipulation reveals, by direct observation of its property, some information on its object [4]. This definition represents a de-trivialization of the concept of icon as a similar entity. If an icon can be characterized as a sign consisting of interrelated parts that reveals information through its manipulation followed by observation [3–5] and, if these relations are subject to experimental modifications regulated by rules, we are working with *diagrams*. Based on those rules, the translation creates a target-sign, in which selected properties from the source are transcreated to the target. Since target and source have different physicalities, the intersemiotic translation recreates the selected characteristics in different materials in a different semiotic system, revealing important and sometimes not easily perceived, semiotic and material perspectives on the process, on the source-sign and the on target-sign.

## 3 The Intersemiotic Translation of the *Fachwerkhäuser des Siegener Industriesgebietes* and the Models of Iconic Semiosis

We approach intersemiotic translation basing on two models [6]: (i) the source of the translation is the sign, and the target is the interpretant (Fig. 2), (ii) the source is the object and the target is the sign (Fig. 3). The first model highlights the production of the target as the interpretative effect of an intersemiotic translation (the target is a cognitive system), and the second one highlights the production of an effect on a cognitive system, that might be a reader of a book, a person that goes in an art exhibition or even an audience composed of several people. By applying the models to the translation from the architectural landscape of the Siegerland Region, Germany (as illustrated in Fig. 1(a)), to the exhibitions *Fachwerkhäuser des Siegener Industriesgebietes* (as illustrated in Fig. 1(b)), we derive the following models:

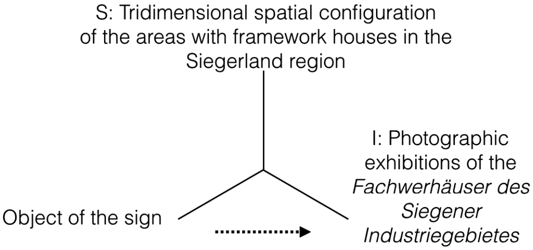


**Fig. 2.** The intersemiotic translation of the *Fachwerkhäuser des Siegener Industriesgebietes* applied to the first model.

According to this model, we have the source (the tridimensional spatial configuration of the areas with framework houses in the Siegerland region) as the sign of the relation, the target (the photographic exhibitions) as the interpretant, and the object of the source as the object of the semiosis. The consequence of approaching the source as the sign instead of as the object, is to stress that the same source has the capacity of determining several different semiotic objects. In this case, we have the architectural pattern revealed by source as the source's object, that produced the exhibition as its interpretant in an intersemiotic translation.

As the second model shows, the source of the intersemiotic translation (the tridimensional spatial configuration of the areas with framework houses in the Siegerland region) is behaving as the object, the target (the photographic exhibitions) as the sign, and the effect that the sign might produce in a potential cognitive system as the interpretant. This model stresses the production of an effect on a cognitive system, and the consequence of it is the creation or revealing of new and/or surprising information that would lead even to the accomplishment of more intersemiotic translation processes or similar processes of aesthetic critic and creation. According to our example, the second model is in dialogue with the affordances and procedures adapted from the Becher's works into the aesthetic principles of the, for example, so called Düsseldorf School of Photography (where the Becher's used to lecture), composed by artists such as Andreas Gursky, Candida Höffer, Thomas Ruff, Thomas Struth and Axel Hütte.

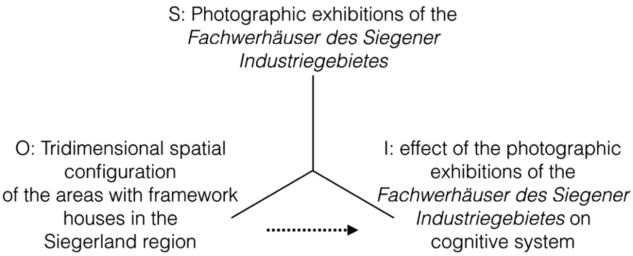


**Fig. 3.** The intersemiotic translation of the *Fachwerkhäuser des Siegener Industriesgebietes* applied to the second model.

## 4 Conclusion

Intersemiotic translation iconically replicates semiotic experimental modifications regulated by rules to observe *how* another sign system produces analogous effects. The operational criterion of iconicity connects *discovery* to the manipulation of diagrams as a process of translation in which the physicality of the source- and target-signs has the ontology of a relation. What is translated by the target is a system of rules and regulations. In this sense, the target reveals a multilevel system of constraints that is analogously observable in the source - in the case of the *Fachwerkhäuser des Siegener Industriesgebietes,* this multilevel system of constraints is a set of architectural-topological properties, that is analogously observable in the tridimensional spatial configuration of the areas with framework houses in the Siegerland region.

## References

1. Jakobson, R.: Selected Writings. Mouton & Co., Paris (1971)
2. Benjamin, W.: Die Aufgabe des Übersetzers. Gesammelte Schriften IV- I, Frankfurt (1923)
3. Stjernfelt, F.: Diagrammatology. Springer, Dordrecht (2007)
4. Peirce, C. In: Hartshorne, C., Weiss, P., Burks, A. (eds.): The Collected Papers of Charles Sanders Peirce. Electronic edition, vol. 8. InteLex Corporation, Charlottesville, VA (1994)
5. Hookway, C.: Truth, Rationality, and Pragmatism. Clarendon Press, Oxford (2002)
6. Queiroz, J., Aguiar, D.: C.S. Peirce and intersemiotic translation. In: Trifonas, P. (ed.) International Handbook of Semiotics, pp. 201–215. Springer, Dordrecht (2015)

# Author Index